Changes in statistical models : various approaches in automatic control and statistics

Michèle Basseville

IRIA

CENTRE DE RENNES

IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: 954 90 20

Rapports de Recherche

N° 72

CHANGES IN STATISTICAL MODELS: VARIOUS APPROACHES IN AUTOMATIC CONTROL AND STATISTICS

Michèle BASSEVILLE

Mai 1981

CHANGES IN STATISTICAL MODELS: VARIOUS APPROACHES IN AUTOMATIC CONTROL AND STATISTICS

Michèle BASSEVILLE
IRISA/CNRS

Résumé :

Le but de cet article est de présenter un panorama des diverses approches du problème des ruptures de modèles statistiques, telles qu'elles sont apparues ces dernières années dans la littérature relevant des domaines de l'Automatique et de la Statistique. L'originalité de ce travail se situe dans une tentative de confrontation des méthodes utilisées en Automatique et en Statistique.

Les diverses méthodes recensées sont classées selon la complexité des modèles qu'elles permettent d'étudier, et selon leur "type" (approches bayésiennes, points de vue séquentiels ou non séquentiels, analyses par blocs glissants...).

Enfin, on présente quelques études comparatives. L'étude d'un exemple réel de ce problème est décrite dans le deuxième article de ce rapport:
"UN EXEMPLE DE DETECTION DE RUPTURES : ELABORATION ET ETUDE COMPARATIVE DE QUELQUES ALGORITHMES".
par M. BASSEVILLE et A. BENVENISTE.

Abstract :

This survey paper is devoted to the problem of changes in statistical models as it has been investigated in the literature concerning Automatic Control as well as Statistics. The originality of this paper is a tentative confrontation of the methods used in these two domains.

The various methods are presented and ordered according to the complexity of the models they allow to be analyzed and according to their "type" (bayesian appraoches, sequential or nonsequential framework, analysis by sliding blocks...).

Finally, some comparative studies are presented. An analysis of a real example of this problem is reported in the second paper of this report:
"AN EXAMPLE OF FAILURE DETECTION: DESIGN AND COMPARATIVE STUDY OF SOME ALGORITHMS"
by M. BASSEVILLE and A. BENVENISTE.

CHANGES IN STATISTICAL MODELS: VARIOUS APPROACHES IN AUTOMATIC CONTROL AND STATISTICS

M. BASSEVILLE

INTRODUCTION.

The purpose of this paper is to present an overview of the different approaches for the problem of changes in statistical models, as they have appeared during the last years in the literature concerning Automatic Control as well as Statistics.

This problem has already been investigated in [2], [3b], [17b], [48], [54]. However, the references concerning Automatic Control which are examined in [54] are prior to 1976; and, on the other hand, the more recent studies [17b] and [48] are concerned with Statistics and Econometry. Therefore, the originality of this work, as it has been sketched in [2] and [3b], seems to lay in a tentative confrontation of the methods used in Automatic Control and in Statistics.

Let us first emphasize that the purely "technical" approach of the problem, as it can appear in the design of failure detectors for electronic circuits for example, will not be considered at all; for this type of approach, the interested reader is refered to [40].

After a short presentation of the general problem of the study of changes in statistical models, of the main questions of interest, and of some examples, the different investigated methods will be presented and ordered according to the complexity of the models they allow to be analyzed, and, as far as possible, according to their "type" (sequential probability ratio tests, cumulative sums tests, estimation by maximum a posteriori probability, etc...); this last distinction being sometime tricky (the recursive formulas for a sequential probability ratio test leading to a "cusum" type test), the somewhat arbitrary feature of this classification has to be noticed.

Finally, some comparative studies, which will consider only part of the numerous methods introduced, will be presented.

I. FORMULATION OF THE PROBLEM.

Some key concepts introduced by A.S. WILLSKY in [54] will be resumed in this section.

The study of changes in statistical models is concerned with the following problem :

Let us observe a signal or a dynamical system, the most often described by a linear stochastic model, some characteristics of which may be subjected to sudden changes or "failures", and examine, for each of these failures, the following three tasks:

i) detection of the "failure",

ii) estimation of its magnitude,

iii) possibly, updating the fitting procedure of the system after the failure.

According to the practical cases, only the answer to the first or the two first questions can be of interest, because the requirements of such an analysis are different for a sequential method of fitting or when a failure is looked for somewhere in a fixed size sample. Furthermore, the difficulty in solving these problems is not the same if the characteristics of the system before and after the jump are assumed to be known and the only unknown variable is the failure time (which is often the case for the asymptotic studies proposed by Statisticians), or if all the characteristics before and after the jump are unknown. From a sequential point of view, a reasonable framework seems to be the following: the characteristics of the system before the failure are assumed to be known (and, if not, are identified), but are unknown the characteristics after the failure, and also the failure time and the magnitude of the failure. This point will be further investigated in the sequel.

This problem of segmentation of signals, which can also be considered as a first possible approach to the identification of nonstationary signals, raises in various domains of application. For example, let us mention:

- the aerospatial domain, where the complexity of the controlled systems compells using numerical techniques for failure detection in captors ([8], [16], [33], [34]);

- medicine, with the automatic segmentation of electroencephalograms [31], or the automatic study of electrocardiograms for detecting arrythmias in the heartbeats ([25a],[25b]);
- speech processing: segmentation of the signal for speech recognition purposes ([35a], [35b]), or detection of transmission errors for encoded signals, using simple analysis of the signal rebuilt at the receivor ([49], [50]);
- image processing, where the edge detection problem can be solved by the detection, line by line, of jumps in the mean grey level [3a];
- digital transmission systems in which sudden changes in the characteristics of the channel (for example, phase hits [19]) may occur;
- econometry, where forecasting studies have to be combined with the study of changes in regression models ([38], [48]);
- geophysics, where the detection of transitions between various geological layers is of interest [3c];
- and even detection of incidents on freeways [58].

Let us emphasize that, in all these examples, the word "failure" has not necessarily a physical meaning as in the case of failures in captors aboard airplanes. However, it will be kept in all the sequel, with the general meaning of abrupt change in one or several parameters characterizing a model which describes the observed signal or system. The words "change" and "jump" will be also indifferently used.

In order to solve the three problems related to the study of changes in models which have been mentioned earlier, algorithms have to be designed, which satisfy two tradeoffs, closely related to each other:

a/ the failure detector has to be able to give the alarm as quickly as possible after a failure (minimum time delay for detection) without giving raise to too many false alarms during normal system operation (high mean time between false alarms). These two needs are obviously contradictory, at least because of the sensitivity to high frequencies that the ability of quick detection of sudden jumps involves. The way for solving this tradeoff depends

upon the considered application: in speech processing for example, non detections are greater nuisances than over-segmentation; on the contrary, in case where the designed analysis system involves, at each alarm, a processing with the aid of a complex, time consuming and memory fulling algorithm, one has to be more careful with the occurence of false alarms.

b/ On the other hand, and this second tradeoff refers to the remark which has just been mentioned, for practical reasons it is often necessary to evaluate the efficiency of the failure detection algorithm by taking into account its complexity, and to find a tradeoff between these two criterias. Furthermore, it has to be noticed that the complexity of a failure detection system is not only of computational character but also of technological character: some procedures require the simultaneous presence of several identical captors and explicitly use the redundancy in the information for failure detection (that is often the case in aerospatial domain; see [16], [34] for example), and reducing such a complexity without degrading the performances of the detector may be of interest.

The following sections are devoted to the presentation of the different algorithms which have been proposed for the study of failures in models, the complexity of which goes increasing, from simple changes in mean to the identification of time-variable non-linear systems, through jumps in one or several coefficients in ARMA models, for scalar or vectorial signals. But, except for one example, the vectorial case will not be mentioned before Section V.

II. CHANGES IN MEAN.

A bibliographical study of this problem already appeared in [2] and [3b]. The purpose of this section is to complete it, in a non exhaustive manner, of course. Let us emphasize that the algorithms which will be presented in the following sections (for regression and ARMA models) may obviously be used for the detection of changes in mean, and will not be presented from this point of view in the present section.

One is here concerned with the study of changes in mean in a sequence of independent observations of a scalar signal which is supposed to be gaussian, if it is real valued, or to be a Bernoulli trial, if it is $\{0,1\}$ valued.

More precisely, let $Y_1,\ldots,Y_n,\ldots$ be a sequence of independent observations of a process governed by the probability law:

$$P_0 = N(\mu_0,\sigma^2) \qquad (\text{resp. } B(p_0)) \quad \text{for} \quad 1\leq n\leq r$$

and

$$P_1 = N(\mu_1,\sigma^2) \qquad (\text{resp. } B(p_1)) \quad \text{for} \quad n\geq r+1,$$

where the failure time r is unknown, and possibly one or several parameters μ_0, μ_1, σ^2 (resp. p_0, p_1) are also unknown. The considered test problem is related to the presence of a failure of jump and challenge the two hypothesis:

H_0 : "*the observations* $(Y_i)_{1\leq i\leq n}$ *are governed by the law* P_0" (i.e. $n\leq r$).

and

H_1 : "*the observations are governed by the law* P_0 *for* $1\leq i\leq r$ *and the law* P_1 *for* $r+1\leq i\leq n$" (i.e. $n\geq r+1$),

where n is the current time in a sequential framework, or the fixed size N of the considered sample in a nonsequential framework.

The introduced test statistics often occur to allow, more or less directly, an estimation of the failure time at least (for example, the estimate of the failure time is one stopping-time for the statistics), and therefore some authors will appear both in the item "estimation" and the item "test" of the summary table which will be found at the end of this section.

1/ Estimation problem.

• The maximum likelihood estimates $\hat{r}$, $\hat{\mu}_o$, $\hat{\mu}_1$ (and possibly $\hat{\sigma}^2$) or $\hat{r}$, $\hat{p}_o$, $\hat{p}_1$ have been analyzed by D.V. HINKLEY (see [3b]) who built up an efficient method for a numerical approximation of the asymptotic distribution of $\hat{r}$.

• From a bayesian point of view, A.F.S. LEE and S.M. HEGHINIAN [36] recently obtained the marginal and joint a posteriori probability laws of the failure time and the magnitude of the failure, when the characteristics of the system, which are all unknown, are governed by the following a priori laws:

- gaussian zero-mean laws, with different variances, for the mean of the observations before the jump and for the magnitude of the jump;
- uniform law (over the time interval during which the observations are taken) for the failure time;
- "uniform" law over $\mathbb{R}_+^*$ for the standard direction σ; all these parameters being assumed stochastically independent from each other.

From these laws, some estimates can be deduced: for example, estimates by a posteriori mode.

2/ Test problem.

a) Nonsequential framework.

• In the gaussian case, D.M. HAWKINS [26] derived the distribution, under H_o, of the likelihood ratio statistics whether σ is known or not. In this later case, a correction has been brought up by K.J. WORSLEY [60]. On the other hand, [2] has already provided the interested reader with the results proposed by D.V. HINKLEY who used exponential approximations for the distributions of maxima of random walks and so derived more explicit asymtotic indications (r and N-r infinite). It has to be noticed that, from this asymptotic point of view, it does not matter to assume that μ_o and μ_1 are known or not, since the asymptotic distributions of the estimates and the test statistics are the same in the two cases.

• Still in a nonsequential framework, J. DESHAYES and D. PICARD [17a] studied, in this particular case of jump between two known means, the asymptotic behaviour (in terms of exponential speed of the level and the power) of two tests based upon the cumulative sums of the innovations or of their squares, tests which have been proposed by R.L. BROWN, J. DURBIN, J.M. EVANS [7]. These asymptotic results were obtained with the aid of results concerning large deviations for random walks.

• In order to study jumps in the parameter of a Bernoulli law in a sequence of independent observations, A.N. PETTITT [43] introduced another test statistics, also of the "cusum" type, but in a nonsequential framework. Let be: $S_N = \sum_{i=1}^{N} Y_i$, where N is the size of the observed sample. The considered statistics is of the form:

$$U_n = N S_n - n S_N$$

or, equivalently,

$$U_n = N \sum_{i=1}^{n} (Y_i - \frac{S_N}{N}).$$

Let: $p_o = P(Y_i=+1) = 1 - P(Y_i=0)$ for $i \leq r$

and $p_1 = P(Y_i=+1) = 1 - P(Y_i=0)$ for $i \geq r+1$

be the state probabilities before and after the jump. Then, when $p_o > p_1$, the failure time can be estimated by the time at which U reaches its maximum value, i.e.:

$$\tilde{r} = \inf \{n \mid U_n \geq U_k, \quad k=1,\ldots,N\}$$

For testing the hypothesis H_o againts the hypothesis H_1, the following test statistics may be used:

$$V^+ = \max_{1 \leq n \leq N} U_n,$$

$$V^- = \min_{1 \leq n \leq N} U_n,$$

or:

$$V = \max_{1 \leq n \leq N} |U_n|$$

according to the a priori knowledge of the "direction" of the jump ($p_o > p_1$ or $p_1 > p_o$). A.N. PETTITT showed that, since S_N is a sufficient statistics for p_o,

the distributions, under H_o, of V^+, V^-, V conditionally to $S_N=m$ are the same as the distributions, under H_o, of $m(N-m)\ D^*_{m,N-m}$, where $D^*_{m,N-m}$ is the corresponding Kolmogorov-Smirnov test statistics (* means +, - or no upperscript).

A.N. PETTITT compared these statistics to the likelihood ratio:

$$W_n = \sum_{i=1}^{n} \left[Y_i \text{ Log } \frac{\overline{Y}_n}{\overline{Y}_{n^*}} + (1-Y_i) \text{ Log } \frac{1-\overline{Y}_n}{1-\overline{Y}_{n^*}} \right] + \sum_{i=1}^{N} \left[Y_i \text{ Log } \overline{Y}_{n^*} + (1-Y_i) \text{ Log}(1-\overline{Y}_{n^*}) \right],$$

where: $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\overline{Y}_{n^*} = \frac{1}{N-n} \sum_{i=n+1}^{N} Y_i$,

which is involved in test statistics of H_o against H_1; for example:

$$W_N - \max_{1 \le n \le N} W_n, \quad \text{or} \quad W_N - W_{\hat{r}},$$

where $\hat{r}$ is the maximum likelihood estimate of r, as D.V. Hinkley proposed (see the next paragraph). A simulation study showed that the two tests, based upon W_n and upon V, have nearly the same power for a fixed level, and that, when $r=\frac{N}{2}$, the test based upon V is more powerful. Since the distribution, under H_o, of $W_N - \max_{1 \le n \le N} W_n$ is furthermore incalculable, thus the test based upon V seems to be preferable.

Then, A.N. PETTITT considered the problem of estimation of r with the aid of the statistics U_n, and compared, from an asymptotic point of view as well as for small samples, this estimate to the maximum likelihood one. This point will be further investigated in section VII, which is devoted to comparative studies.

Finally, the same author proposed a sequential procedure, which, because of its complexity, seems to be unrealistic; it proceeds in the following manner when several failures possibly occur: test H_o against H_1 using the first k observations (k=2,3,...,N); then examine the probability p_k that no jump occured and increase k until this probability reaches a local minimum less than a threshold (of order of magnitude 0.1 or 0.05); this minimum at k_o indicates a possible failure, the occurence time of which can be evaluated with the aid of the estimate $\tilde{r}$ which has been mentioned earlier and the first k_o observations. After time (k_o+1), the observations are again analyzed in the same way. No example is given.

b) Sequential framework.

• Tests and methods of detection based upon cumulative sums ("cusum" tests) are frequently used. M. BASSEVILLE investigated in [3b] one of these algorithms which is based upon the deviations of these cumulative sums with respect to their maxima, and which was proposed by E.S. PAGE and D.V. HINKLEY. It proceeds in the following way: when μ_o and μ_1 are known, a change in the sign of the mean has in fact to be tested. So, define: $S_o=0$ and $S_o=0$ and $S_n = \sum_{i=1}^{n} (Y_i - \frac{\mu_o+\mu_1}{2})$ $(n\geq 1)$. If $\mu_o<0<\mu_1$, $m_n = \min_{o\leq k\leq n} S_k$ is computed, and the alarm is given as soon as: $S_n-m_n \geq h$, where h is a fixed threshold. If $\mu_o>0>\mu_1$, $M_n = \max_{o\leq k\leq n} S_k$ is computed, and the alarm is given as soon as: $M_n-S_n \geq h$; see *figure n° 1*. When μ_o and μ_1 are unknown, a reasonable procedure consists in identifying μ_o and fixing a minimum magnitude of jump: $\nu = |\mu_1-\mu_o|$ to be detected. Two tests are then activated in parallel:

on one hand:

$$\begin{cases} S_n = \sum_{i=1}^{n} (Y_i - \mu_o - \frac{\nu}{2}) \quad (n\geq 1) \quad (S_o=0) \\ m_n = \min_{o\leq k\leq n} S_k \\ \text{alarm when } S_n-m_n \geq h \end{cases}$$

is looking for upwards jumps ($\mu_o \rightarrow \mu_o+\nu$), whereas, on the other hand:

$$\begin{cases} T_n = \sum_{i=1}^{n} (Y_i - \mu_o + \frac{\nu}{2}) \quad (n\geq 1) \quad (T_o=0) \\ M_n = \max_{o\leq k\leq n} T_k \\ \text{alarm when } M_n-T_n \geq h \end{cases}$$

is looking for downwards jumps ($\mu_o \rightarrow \mu_o-\nu$).

Then, there are two possible estimates for the failure time: either the alarm time, possibly reduced by its "bias" (i.e. mean time delay for detection of a jump, conditionally to the fact that no false alarm occurs before the jump), or the instant at which the maximum of T_k (resp. the minimum of S_k) has been reached.

This detector has been compared, via a simulation study, to some others, of the "filtered derivatives" type which will be presented later, and to a A.N. SHIRYAEV's algorithm (sequential probability ratio test; see *figure n° 2*). Furthermore, new theoretical results, concerning mean time between false alarms and mean time delay for detection of a jump, have been derived in the two cases considered here: change in mean for a gaussian law, change in the parameter of a Bernoulli law. See [3b].

It should be noticed that the setting of the threshold h, proportional to the variance σ^2 of the observations (and not the standard deviation σ), requires estimation of this variance when unknown.

• In the gaussian case, W.G.S. HINES ([27], [28]) proposed a weighted version of this cusum technique, which he called "geometric moving average", and can be described in the following manner: let s_n be the state of the system at time n (i.e. $s_n=j$ if and only if $Y_n \sim N(\mu_j,\sigma_j^2)$, where $j=0,1$), and let $\hat{s}_n$ be the estimate of this state. The considered statistics has the form:

$$G_n = Y_n + \alpha(\hat{s}_{n-1})\, G_{n-1},$$

where α is a weighting coefficient which possibly depends upon the estimated state $\hat{s}_{n-1}$. The alarm is given as soon as: $G_n \geq h(\hat{s}_{n-1})$, where h is a threshold which also possibly depends upon $\hat{s}_{n-1}$, and the estimated state becomes: $\hat{s}_n = 1 - \hat{s}_{n-1}$, and α and h are modified if needed, whereas the test statistics is updated to the value:

$$G_n = \frac{\mu_{\hat{s}_n}}{1 - \alpha(\hat{s}_n)}$$

which is the mean value of the distribution of G_n when the system is always under the state $\hat{s}_n$. See *figure n° 3*.

By approximating the sequence $(G_n)_n$ with a Ornstein-Uhlenbeck's process and using approximated distributions of level-crossing times for such a process, W.G.S. HINES obtained evaluations of the efficiency of such a failure detector, in terms of linear combination of frequencies of the two possible errors which can occur during estimation of the state of the system, under the hypothesis that actual failures as well as false alarms scarcely occur.

Such a measure of efficiency allowed him to set the various parameters of his detector (α and h) for each system. It should be noticed that the knowledge of the means and variances before and after the jump is required. Under the hypothesis that the variance of the observations is equal to one and that their mean varies from $-\mu$ to $+\mu$, a simulation study led to an evaluation of the precision of the involved approximations, and of the efficiency of the proposed detector with respect to the bayesian one:

$$R_n = e^{2\mu Y_n} \frac{(1-p) R_{n-1} + p}{1-p + p R_{n-1}} ,$$

which is optimal for the criteria "mean time in error". (p is the a priori probability for the system to be in the state 0).

In [28], W.G.S. HINES proposed an improved version of this "geometric moving average" detector, for which the measure of efficiency takes into account the frequency of the false alarms, and the updating of the statistics G_n is modified: after an alarm, G_n is updated to the value:

$$\lambda(\hat{s}_n)\ h(\hat{s}_n) + [\,1 - \lambda(\hat{s}_n)] \frac{\mu_{\hat{s}_n}}{1 - \alpha(\hat{s}_n)} ,$$

which lies between the new threshold $h(\hat{s}_n)$ and the mean value of the distribution of G_n when the system is always in the state $\hat{s}_n$. A numerical study showed that $\lambda = \lambda(o) = \lambda(1)$ has to decrease when the cost of false alarms increases.

• The main drawback of all these parametric tests is that they require a preliminary identification of the model. Thus, let us mention some non-parametric tests (or nearly non-parametric tests) which may be useful. As far as linear rank tests are concerned, the interested reader is refered to [22] and [35a] (chp.4). We shall only describe here "filtered derivatives" detectors, which have been mentioned earlier in this section and studied, for instance, in [3b], [3c].

These simple well known detectors first filter the analyzed signal before comparing its derivative to a threshold λ. Two simple filters have been considered in [3b]: the integrating filter and a filter, called "triangular", the impulse response of which is of the form:

$$- a\,t\,1_{\{o\leq t\leq \ell\}} - a(2\ell - t)\,1_{\{\ell\leq t\leq 2\ell\}}, \quad \text{with} \quad a\ell^2 = 1.$$

After one of these filtering operationsand derivation (in discrete time), the obtained observations are:

$$Z_n = \frac{Y_{n+\ell} - Y_{n-\ell}}{2\ell}$$

in the first case, and:

$$Z_n = \frac{(Y_{n+1}+Y_{n+2}+\ldots+Y_{n+\ell}) - (Y_{n-1}+Y_{n-2}+\ldots+Y_{n-\ell})}{\ell^2}$$

in the second.

With the aid of these two detectors, two ways of setting the alarms are possible: the first one, called "rough" consists in giving the alarm when $|Z_n| \geq \lambda$ and forgetting the 2ℓ observations following this instant, in order to avoid successive detections in the neighborhood of a jump. The second method, called "with counter", consists in counting the number of instants at which the new signal $|Z_n|$ exceeds the threshold λ, and which occur within the 2ℓ instants following the first overcrossing, and in setting the alarm only if this number of crossings exceeds a fixed value N_c; the alarm time is then chosen to be the first of these crossing-times. The advantage of such a method is obviously to avoid false alarms, the drawback being the risk of nondetection if N_c is too high. It was shown that $N_c=2$ was a good value, in the simulation analysis of simple jumps in mean [3b] as well as in the framework of edge detection in noisy pictures [3a].

However, when applied to very noisy signals, these detectors seem to be inefficient and a more elaborated type of counter has been studied in [3c]: this new counter takes into account Z_n itself and not only its absolute value, and the alarm is now given when the number of crossings of the same type (i.e. overcrossings of the level $+\lambda$ or undercrossings of the level $-\lambda$) exceeds the threshold N_c during the 2ℓ instants following the first one, which is taken as the estimated jump time.

• To complete this section, devoted to changes in mean, let us mention a particular case which has been investigated by S. ZACKS [62] in his numerical study of distributions of some stopping-times which are used in sequential

failure detection: the case of a change in the parameter of a Poisson distribution. This point will be further analyzed in section VI which is devoted to changes in "general" models.

Summary table.

	Nonsequential			Sequential	
	Bayesian	Max. Likelihood	"Cusum"	Bayesian	"Cusum"
Estimation	Lee et al.	Hinkley	Hinkley		Hinkley, Basseville Pettitt
Test		Hawkins Worsley Hinkley	Hinkley Deshayes-Picard Pettitt	Zacks	Hinkley, Basseville Hines Zacks Pettitt

III. CHANGES IN LINEAR REGRESSION MODELS.

A great number of bibliographical references concerning this problem will be found in [2], [3b], [17b], and an annotated bibliography is given in [48] where problems encountered in Econometry are emphasized. On the other hand, the reader is referred to section VII for comparative studies of failure detectors in regression models.

Only some recent approaches will be presented here. As in the previous section, many approaches used for more complex models which will be presented in the next sections, may be used in the present framework of regression models.

Let : $y_n = x'_n \theta_n + \varepsilon_n$

be the studied regression model, where $\theta_n \in \mathbb{R}^m$ is the parameter vector, $x_n \in \mathbb{R}^m$ is a vector of non random variables, ("regression" variables), and the ε_n are a sequence of independent random variables, gaussian, zero-mean, with constant variance σ^2. The hypothesis H_o of absence of failure can be written:

$$\theta_1 = \theta_2 = \ldots = \theta_n = \theta.$$

1/ Nonsequential estimation.

We recall that, from a nonsequential point of view, the problems of estimation of the parameters θ and of the failure time r with maximum likelihood have been investigated by D.V. HINKLEY, and that the asymptotic behaviour of the likelihood ratio when several failures may occur has been analyzed in detail by P.I. FEDER. The interested reader is referred to [3b] for precise references.

2/ Sequential test and estimation.

From a sequential point of view, when the "regression" variable x is the time itself, test and estimation problems can be approached by using the Hinkley's "cusum" statistics which hasbeen presented in section II.2/*b)*:

one has only to use a Kalman filter adapted to the studied regression model, and to apply the sequential detector of change in mean to the innovations of this filter, as it has already been done in [3a] and [22] for the line-by-line detection of edge elements in digitized pictures.

3/ Some nonsequential tests.

The remainder of this section is devoted to some nonsequential approaches of the test problem for the constancy of the regression parameters θ.

• B.P.M. Mac CABE and M.J. HARRISON [38] introduced a test based upon least squares residuals, and compared it to the test proposed by R.L. BROWN, J. DURBIN, J.M. EVANS [7]. It has to be noticed that the authors themselves think of this test more as an aid for data analysis than as a precise test. Let E be the vector of the N residuals coming from least squares estimation of θ:

$$E = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix} = (I_N - X(X'X)^{-1} X')Y,$$

where $Y = (y_1,\ldots,y_N)'$, and $X = (x_1,\ldots,x_N)'$.

The test consists in observing the statistics:

$$s_n = \frac{E' A_n E}{E' E},$$

where $A_n = \begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix}$, as a function of n, and its crossing-times of the levels:

$$\pm s_u^* = \pm c + \frac{n}{N-m} \quad \text{for} \quad 1 \leq n \leq N-m-1,$$

and : $$\pm s_L^* = \pm c + \frac{n-m}{N-m} \quad \text{for} \quad m+1 \leq n \leq N-1.$$

(see *figure n°4*).

For example, if the hypothesis H_1 is:

"$\theta_n = \theta$ for $1 \leq n \leq r$ and $\theta_n = \theta + \Delta$ for $r+1 \leq n \leq N$",

then the cumulative sums of squares s_n tend to lay under the straight line

joining the points $\frac{n}{N}$ $(1 \leq n \leq N)$, and the test of H_o against H_1 uses the levels $-s_U^*$ and $-s_L^*$, where c is chosen in order to ensure a given level. (This choice is made with the aid of a brownian motion approximation). This test statistics is easier to compute than this of R.L. BROWN, J. DURBIN, J.M. EVANS [7] which cumulates the squares of the innovations weighted by their variances:

$$T_n = E' B_n E$$

where $B_n = \begin{pmatrix} V_n & 0 \\ 0 & 0 \end{pmatrix}$ and where V_n is diagonal, of i^{th} element $\frac{1}{\text{var}(e_i)}$ (which is estimated together with θ). Therefore, B.P.M. Mac CABE and M.J. HARRISON compared, via a simulation study, the powers of these two tests by using the simulation method proposed by K. GARBADE [21] (see section VII) for samples of rather small sizes (N from 15 to 60). No noticeable difference between the two powers appeared. However, if the sample size is very small and if the failure is of small magnitude, BROWN et al.'s test seems to be more powerful. Finally, B.P.M. Mac CABE and M.J. HARRISON give an example of application of their test to econometric forecasting models for telephone traffic.

• Let us now examine the asymptotic results recently derived by P.K. SEN [47], which are related to two test statistics for changes in regression models, one of them being nonparametric. (The reader is referred to [3b], [17b], [22] and [48] for other references of such tests).

Let $y_i = y(t_i)$ $(1 \leq i \leq N)$ $(t_1 \leq \ldots \leq t_N)$ be the observations, and consider the following regression model:

$$y_t = \begin{cases} \alpha + \beta(t-r) + \varepsilon_t & \text{if} \quad t \leq r \\ \alpha + \gamma(t-r) + \varepsilon_t & \text{if} \quad t > r \end{cases},$$

where α, β, γ and r are unknown, and where the ε_t are independent, zero-mean random variables, with distribution function F. For testing the hypothesis H_o: $\beta=\gamma$ against the hypothesis H_1: $\beta>\gamma$ or the hypothesis H_2: $\beta \neq \gamma$, P.K. SEN proposes tests which are based upon least squares estimates and a linear rank statistics, the asymptotic properties of which are studied. The asymptotic comparison between these two tests will be further investigated in section VII.

The tests based upon the least squares estimates are built in the following manner: let be:

$$\bar{t}_n = \frac{1}{n} \sum_{i=1}^{n} t_i \qquad \text{and} \qquad T_n^2 = \sum_{i=1}^{n} (t_i - \bar{t}_n)^2 \quad (n \geq 1).$$

Under H_o, the least squares estimate of β based upon the observations $y_1,\ldots,y_n$ is of the form:

$$\hat{\beta}_n = \frac{1}{T_n^2} \sum_{i=1}^{n} (t_i - \bar{t}_n) y_i \qquad \text{for} \quad 2 \leq n \leq N$$

and $\hat{\beta}_1 = 0$. Let $\hat{y}_n = y_n - \hat{\beta}_N t_n$ $(1 \leq n \leq N)$ be the residuals, and let us compute:

$$\tilde{\beta}_n = \frac{1}{T_n^2} \sum_{i=1}^{n} (t_i - \bar{t}_n) \hat{y}_i$$

$$= \hat{\beta}_n - \hat{\beta}_N \qquad (2 \leq n \leq N).$$

The considered test statistics are:

$$M_N^+ = \max_{0 \leq n \leq N} S_{N,m}$$

and:

$$M_N = \max_{0 \leq n \leq N} |S_{N,n}|,$$

where $S_{N,0} = S_{N,1} = S_{N,N} = 0$ and:

$$S_{N,n} = \frac{T_n^2}{T_N} \tilde{\beta}_n$$

$$= \frac{T_n^2}{T_N} (\hat{\beta}_n - \hat{\beta}_N) \qquad (2 \leq n \leq N-1),$$

and the test procedure is the following one: when F has a known finite variance σ^2, reject H_0 if M_N^+ or M_N exceeds the threshold $\sigma \Delta_\varepsilon^+$ or $\sigma \Delta_\varepsilon$, where Δ_ε^+ and Δ_ε are the respective roots of:

$$1 - e^{-2x^2} = 1-\varepsilon$$

and:

$$1 - 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2} = 1-\varepsilon,$$

ε being the desired level. (These equations are obtained when $S_{N,n}$ is approximated by a brownian bridge). When σ^2 is unknown, it is estimated by:

$$\hat{\sigma}_N^2 = \frac{1}{N} \sum_{n=1}^{N} (y_n - \overline{y_N} - \hat{\beta}_N (t_n - \overline{t_N}))^2, \quad \text{where} \quad \overline{y_N} = \frac{1}{N} \sum_{n=1}^{N} y_n,$$

and the same procedure is used, with σ being replaced by $\hat{\sigma}_N$.

The nonparametric method P.K. SEN considers is based upon linear rank tests, which are obtained from the following statistics. Let $R_{nk}(b)$ be the rank of $y_k - bt_k$ in the sequence $y_1-bt_1,\ldots,y_n-bt_n$ $(1\leq k\leq n)$, and let us consider the linear rank statistics:

$$L_n(b) = \sum_{k=1}^{n} (t_k-\overline{t_n})\, a_n(R_{nk}(b)) \quad (n\geq 1),$$

where $a_n(k) = E[\phi(U_{nk})]$ $(1\leq k\leq n)$, with: $U_{n1} <\ldots< U_{nn}$ an ordered sample of size n of an uniform law over $]0,1[$, and where ϕ is a growing function square-integrable on $]0,1[$.

Then let the estimates be:

$$\beta^*_{n,1} = \sup\{b \mid L_n(b)>0\},$$

$$\beta^*_{n,2} = \inf\{b \mid L_n(b)<0\},$$

and $$\beta^*_n = \frac{\beta^*_{n,1} + \beta^*_{n,2}}{2} \quad (1\leq n\leq N),$$

and let us introduce the linear rank statistics related to the later one:

$$L^*_{N,n} = \frac{L_n(\beta^*_N)}{T_N}$$

$$= \frac{1}{N} \sum_{k=1}^{n} (t_k-\overline{t_n})\, a_k\,(R^*_{nk}) \quad (1\leq n\leq N)$$

where R^*_{nk} is the rank of $y^*_k = y_k - \beta^*_N t_k$ in the sequence $y^*_1,\ldots,y^*_n$ $(1\leq k\leq n)$ $(n\geq 1)$. The test statistics proposed by P.K. SEN have the following form:

$$D^+_N = A_N^{-1} \left(\max_{0\leq n\leq N} L^*_{N,n}\right)$$

and:

$$D_N = A_N^{-1} \left(\max_{0\leq n\leq N} |L^*_{N,n}|\right),$$

where, for $N\geq 2$, $A_N^2 = \frac{1}{N-1} \sum_{k=1}^{N} [a_N(k)-\overline{a_N}]^2$ and $\overline{a_N} = \frac{1}{N} \sum_{k=1}^{N} a_N(k)$.

The test procedure is then: reject H_o if: $D^+_N \geq \Delta_\varepsilon^+$ or if: $D_N \geq \Delta_\varepsilon$, where Δ^+_ε and Δ_ε are the roots of the same equations as before.

For these two types of tests, built upon least squares estimates or linear rank statistics, P.K. SEN studies the asymptotic power for a sequence of local alternative hypothesis:

$$H_n: \quad \beta = \gamma + \frac{\delta}{T_n} ,$$

and derives from it some comparisons which will be presented in section VII.

IV. CHANGES IN AR AND ARMA MODELS.

This section is devoted to the presentation of some techniques for failure detection in autoregressive (AR) or autoregressive and moving average (ARMA) models, still for scalar signals, except for the study of L. TELKSNYS [52c] who investigates the vectorial case. These techniques use the modelling of the observed signal or system by such processes, and generally are distinct from those that will be presented in the next section, where a state space model is explicitly taken into account.

1/ Changes in autoregressive models.

Three types of processing will be distinguished in this paragraph: a sequential algorithm ; a nonsequential framework in which a bayesian approach and some "cusum" type algorithms will be presented; and then two analysis by sliding blocks, involving either distance-measures between spectral densities or order-based-detectors, will be described.

a) Sequential framework.

Let us first emphasize some general ideas underlying sequential analysis of failures, in autoregressive as well as in autoregressive and moving average processes. The main idea consists in, first, filtering the observed signal (y_t) through a known or identified AR or ARMA filter and to look for failures in the residual signal of "innovations" (e_t). If a bias on the model is looked for, a jump in the mean of the e_t can be detected with the aid of any method which was presented in section II. On the other hand, if jumps in other parameters of the model are expected, looking for jumps in the mean of e_t^2, with the aid of the previous techniques, may be of interest. Finally, a more sophisticated technique consists in computing likelihood ratios between various hypothesis H_1 corresponding to the studied jumps and the null hypothesis H_0. These basic ideas will be found all along this section and the following one, which is devoted to techniques based upon state variables models, and should be kept in mind.

Let us thus examine an algorithm, which has been introduced by V.Y. LUMEL'SKI [37] and studied by L.I. BORODKIN and V.V. MOTTL' [6], and which is based upon cumulative sums of squares of "innovations". This procedure is intended to detect changes in any parameter of a scalar autoregressive process, and it shall be seen in section VI that it is nothing but a particular case of a new failure detector which has been proposed by J. SEGEN and A.C. SANDERSON [46] in a more general framework and studied by them in EEG processing [44]. However, let us describe the algorithm in this simple case. Let:

$$y_n = f(Y_n^p, \theta) + \beta\, z_n \quad (n \geq 1) \quad (\theta \in \mathbb{R}^m)$$

be the AR(p) model, where $Y_n^p = (y_{n-1}, \ldots, y_{n-p})'$, z_n is zero-mean white noise with variance 1, and with known $E(z_t^4) = \gamma$; f is a known function ; the (m+1) parameters $(\theta^{(o)}, \beta^{(o)})$ which characterize the model before the jump are supposed to be known, whereas the parameters $(\theta^{(1)}, \beta^{(1)})$ after the jump, which occurs at the unknown time r, are unknown. Let us introduce the quantities:

$$e_n^2 = \left| \frac{y_n - f(Y_n^p, \theta^{(o)})}{\beta(o)} \right|^2 .$$

For $n \leq r$, $E(e_n^2) = 1$. From which we consider the statistics:

$$Z_n = \frac{1}{\sqrt{(\gamma-1)n}} \sum_{k=1}^{n} (e_k^2 - 1),$$

which is zero-mean, and asymptotically governed by the gaussian law $N(0,1)$ for $n \leq r$. If (n-r) is large enough to allow neglecting the influence of the transitory situation of length p which is established after a failure, then:

$$E(Z_n) = \frac{S(n-r)}{\sqrt{(\gamma-1)n}}$$

where $S = E(e_n^2 - 1)$ for $n > r+p$.

Typically, Z_n is in the neighborhood of zero until r and monotically varies after r (see *figure n°5*). In order to estimate the failure time r, the overcrossings of a level by $|Z_n|$ will be observed and, so as not to penalize long observation sequences where r is large, Z_n is updated to zero every T steps (provided that $|Z_n|$ has not overcrossed the threshold),

where T is fixed a priori as a function of the length of the periods during which the observed signal is "stationary".

The failure detection algorithm is thus the following one:

1 estimate $(\theta^{(o)},\beta^{(o)})$ with the aid of the $\ell<T$ first observations;

2 compute the sequences e_n^2 and Z_n;

3 if $|Z_n|\geq h$, go to step 6;
if not, continue;

4 recursively refine the estimates of $(\theta^{(o)},\beta^{(o)})$ with the aid of y_n; increase n by unity; if $n<T$, go back to 2; if not, continue;

5 update n and Z_n to zero; keep the estimates of $(\theta^{(o)},\beta^{(o)})$;

6 set the alarm; keep $\hat{f}=n$; update n to zero and start again from 1.

This algorithm has been successfully tested over synthetical data, for which T=150, ℓ=100, r=250, h=2.5, and which correspond to the following cases:

- jump in the regression coefficient of an AR(1) model;
- jump in the mean in an AR(1) model;
- jump in the variance (β coefficient) in an AR(2) model;
- jumps in the mean, the variance and the four regression parameters of an AR(4) model;
- jumps in the variance and the regression coefficient of an AR(1) model.

The delay for detection is about 10. It has to be noticed that this algorithm seems not to be too much sensitive to the set up of the threshold h, but requires the knowledge of $\gamma = E(z_n^4)$.

However, heuristic considerations allow to suppose that this algorithm is not as efficient as Shiryaev's or Hinkley's one when they are applied to the "residuals" of the whitening AR(p) filter. Indeed, the threshold to which the cumulative sum Z_n is compared is in fact updated every T instant for the present algorithm, from time to time (according to the evolution of the observations) for Shiryaev's algorithm, and every time (because of the current minimum of the cumulative sum) for Hinkley's one. Compare *figures n°1, 2* and *5*. And, furthermore, the noise sensitivity of Shiryaev's and Hinkley's algorithms under H_o is far less than that of the present algorithm, because their drift under H_o is non zero.

b) Nonsequential techniques.

• Let us first consider L. TELKSNYS's bayesian approach ([52a], [52b], [52c]) which uses the maximum of an a posteriori probability law, when the observed system is vectorial and all the parameters, before and after the failure, are unknown, together with the failure time. The originality of this work seems to lay in an artful writing of these a posteriori densities which allows to avoid inverting large matrices: the idea consists in neglecting some factors of the a posteriori law, under the hypothesis that the number of available observations is far greater than the largest order of the model.

• B. CHALMOND [10] investigated the problem of jump in mean for an autoregessive model, and studied the ratio of maximal likelihoods (i.e. the generalized likelihood ratio which will be described in section V in another framework), the behaviour of which is analyzed under H_o and for a known failure time. The threshold used for this statistics may be determined by studying the crossings of two linear barriers (tangent to a parabole) for a brownian motion, as in [7]. A simulation study has been done for an AR(1) process and showed that this test is uniformly more powerful than the test based upon the cumulative sum of the deviations with respect to the model "constant mean in the independant case". This later test has the advantage to be simple and not to require the knowledge of the order of the model.

Finally, B. CHALMOND studied the bayesian and maximum likelihood estimates of the failure time and the magnitude of the failure and all the unknown parameters before and after the failure (mean and regression coefficients). He noticed that maximum likelihood estimation is nothing but a special case of bayesian estimation with the criterion of maximum mode (and not of minimum risk); and so he avoided the need for a nonlinear method of estimation (indeed the likelihood is not a linear function of the AR coefficients). The two estimation methods have been empirically compared for an AR(1) process: the mean bias of the two estimates are comparable, whereas the mean square deviation of the bayesian estimate is systematically less than that of the maximum likelihood estimate.

c) Analysis by sliding blocks.

Two types of processing will be described in this paragraph; the first one is often used in speech processing and EEG analysis and consists in locally adjusting an AR or ARMA model which is then compared to known models with the aid of distance measures between spectral densities; the final choice for the local model results from a syntactic study of the context, and simplifying heuristics or even dynamic programming techniques are used (compare with multiple model method in section V). The second processing is based upon the jumps in the estimated order of the model.

• Distance measures between spectral densities.

N. ISHII, A. IWATA, and N. SUZUMURA [29] studied the segmentation of nonstationary electroencephalograms signals with the aid of different distance measures between two spectral densities: Kullback's information, Kullback's divergence, Bhattacharrya's distance and Chernoff's distance. These distances allow detecting changes in amplitude and frequency, and thus to segmentate a nonstationary signal in intervals where the signal is stationary, for example by using a reference set of such characteristic intervals (see [31]).

More precisely, consider two AR(m) processes with spectral densities having respective power $p_1(\omega)$ and $p_2(\omega)$. Kullback's information of the former with respect to the later is:

$$I(1:2) = \frac{1}{2}\int_{-\pi}^{\pi} \left(\frac{p_1(\omega)}{p_2(\omega)} + \text{Log}\,\frac{p_2(\omega)}{p_1(\omega)} - 1\right) \frac{d\omega}{2\pi} ,$$

or, following A.H. GRAY and J.D. MARKEL [23]:

$$I(1:2) = \frac{1}{2}\left[\left(\frac{\sigma_1}{\sigma_2}\right)^2 \left(1 + \int_{-\pi}^{\pi} \frac{\left|A_2(e^{j\omega}) - A_1(e^{j\omega})\right|^2}{\left|A_1(e^{j\omega})\right|^2} \frac{d\omega}{2\pi}\right) - \text{Log}\left(\frac{\sigma_1}{\sigma_2}\right)^2 - 1\right]$$

where σ_i^2 (i=1,2) is the prediction error variance for model i, and where $A_i(Z) = 1 + \sum_{k=1}^{m} \alpha_k^{(i)} Z^{-k}$ characterizes the AR model i.

Kullback's divergence :

$$J(1:2) = I(1:2) + I(2:1)$$

is of interest because it is a symmetric function of $d = \text{Log } \frac{p_1(\omega)}{p_2(\omega)}$.

In order to detect changes in amplitude for signals which are supposed to have the same mean, N. ISHII et al. [29] compared the respective sensitivities of Kullback's information, Kullback's divergence and Bhattacharrya's distance for non correlated signals. In this special case, these three distances can be written as:

$$I(1:2) = \frac{1}{2}\left(\text{Log } \frac{\sigma_2'^2}{\sigma_1'^2} + \frac{\sigma_1'^2}{\sigma_2'^2} - 1\right),$$

$$J(1:2) = \frac{1}{2}\left(\frac{\sigma_1'^2}{\sigma_2'^2} + \frac{\sigma_2'^2}{\sigma_1'^2} - 2\right),$$

$$B(1:2) = \frac{1}{2}\left[\text{Log } \frac{\sigma_1'^2+\sigma_2'^2}{2} - \frac{1}{2}\text{ Log } \sigma_1'^2 - \frac{1}{2}\text{ Log } \sigma_2'^2\right],$$

where $\sigma_i'^2$ is the variance of the signal i (i=1,2).

The variations of J, I, B as functions of $K = \frac{\sigma_2'^2}{\sigma_1'^2}$ show that J is more sensitive than I with respect to varying amplitudes, and than I is more sensitive than B (parabolic looling curves, very flat for B, more peaked for J).

Chernoff's distance, which can be expressed by:

$$C(1:2;\ \tau) = \frac{1}{2}\text{ Log } \frac{1+\tau(K-1)}{K^{\tau}} \qquad (0 \leq \tau \leq 1)$$

in this special case, is, at its optimum value, comparable to Bhattacharrya's distance.

In order to detect changes in frequency, N. ISHII et al. compared I, J, B by considering only the term due to correlation, i.e. by supposing $\sigma_1'^2 = \sigma_2'^2$:

$$I(1:2) = \frac{1}{2}\text{ Log } \frac{|S_2|}{|S_1|} + \frac{1}{2}\text{ tr } (S_1\ S_2^{-1} - I),$$

$$J(1:2) = \frac{1}{2} \operatorname{tr} (S_1 S_2^{-1} + S_2 S_1^{-1} - 2I),$$

$$B(1:2) = \frac{1}{2} \operatorname{Log} \frac{\left|\frac{S_1+S_2}{2}\right|}{\sqrt{|S_1|\,|S_2|}},$$

where S_i denotes the autocorrelation matrix and $|S_i|$ its determinant. The curves, which give the variations of I, J, B, as functions of $\frac{|S_2|}{|S_1|}$ for various values of $|S_1|$ for an AR(1) model (i.e. $S_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$), show that Kullback's divergence is the most sensitive to changes in frequency. However, as the sensitivity of Kullback's information is of the same order of magnitude and its computing is easier, Kullback's information has been kept by the authors as a distance measure between spectral densities for segmentation of nonstationary signals [31].

This work has to be compared with the studies of A.H. GRAY and J.D. MARKEL [24a], R.M. GRAY and al. [24b], and D. KAZAKOS and P. PAPANTONI-KAZAKOS [32] concerning spectral distances. According to [24a], Kullback's information and divergence are well approximated, especially for their small values, by the mean square distance between the logarithms of the spectra, which can be written:

$$d_2 = \left(\int_{-\pi}^{\pi} \left| \operatorname{Log} \left(\frac{\sigma_1}{\sigma_2}\right)^2 + \operatorname{Log} \frac{|A_2(e^{j\omega})|^2}{|A_1(e^{j\omega})|^2} \right|^2 \frac{d\omega}{2\pi} \right)^{\frac{1}{2}}$$

and which is of interest because it can be efficiently computed with the aid of the cepstral coefficients and therefore without any Fourier transform.

It has to be noticed that N. ISHII et al. [30] also studied the segmentation of EEG signals with the aid of another method due to T. BOHLIN ([4], see also [2]) which will be further investigated in the paragraph devoted to ARMA processes.

• A test based upon automatic order selection.

The underlying idea of this test is the following one: nonstationarities in a signal involve an increase of the estimated order, and therefore transition areas correspond to peaks in the order estimates. The main

interest of S.A. LAKEHAL's failure detector ([35a], [35 b]) is thus the avoidance of threshold. He uses new formulations adapted from least squares algorithms of Burg's type, and the failure test consists in observing the jumps in the order of the AR model as it is estimated with the criteria of H. AKAIKE, G. SSCHWARZ and J. RISSANEN or J. HANNAN. These three failure tests were compared via a simulation study, in terms of purcentages of non-detections and oversegmentations and of repartition function for the delay to detection. Only H. AKAIKE's criterion allows a rate of nondetections less than 10 %; the high oversegmentation it involves is due to the bias it introduces in the estimation of the order. The choice of the order criterion for the failure test will depend upon the considered application (greater tolerance for nondetections than for false alarms, or conversely).

The practical use of this algorithm, for which the absence of threshold is a priori seducing, seems to require extra studies. Furthermore, non-using of a threshold may be challanged, because the accuracy of the estimate of the order is directly related to the more or less oscillatory feature of the signal.

2/ Changes in ARMA models.

This paragraph is devoted to the presentation of some failure detection algorithms for ARMA(p,q) models (autoregressive of order p, and moving average of order q models). The case q=0 obviously allows to lengthen the list of the previous paragraph.

Let us recall that these algorithms only process scalar signals or systems.

a) Some "cusum" type methods.

• M. BAGSHAN and R.A. JOHNSON [1] proposed an extension of the statistics introduced by E.S. PAGE and D.V. HINKLEY (see sections II.2/*b)* and III.2/) for the case of an ARMA model: their test consists in observing the deviations with respect to its maximum or minimum value according to the "direction" of

the failure, of the logarithm of the likelihood ratio between hypothesis H_o: "the parameter vector θ is θ_o" and H'_1: "the parameter vector θ is θ_1". (Let us notice that D.V. HINKLEY was testing H_o agains H_1: "the parameter θ is θ_o until time r, and θ_1 afterwards"). The test has then the following form:

$$T_n(\theta_0,\theta_1) - \min_{1\leq k\leq n} T_k(\theta_0,\theta_1) \geq h,$$

where:

$$T_n(\theta_0,\theta_1) = \sum_{k=1}^{n} \left[e_k^{(0)2} - e_k^{(1)2} \right]$$

and $e_k^{(i)}$ is the one-step ahead prediction error at time (k-1) due to the use of the parameter θ_i (i=0,1).

The threshold h is chosen with the aid of an approximation by a brownian motion, the drift and the diffusion coefficient of which are given, for instance, as functions of the parameters of an ARMA(1,1) model.

• I.V. NIKIFOROV [42] used an approach of the same type ("cusum" type statistics) for solving the problem of change in mean for an ARMA(p,q) process. His method consists in an extension of E.S. PAGE's and D.V. HINKLEY's test for this case. This extension, which naturally involves the likelihood ratio , has a rather complicated expression, and I.V. NIKIFOROV thus looked for approximations leading to evaluations of the mean time between false alarms and the mean time delay for detection of a jump, and to measures of the sensitivity of the test to a poor knowledge of the mean and the variance of the observations after the jump. Finally, the validity of these approximations has been tested via a simulation study.

b) A nonparametric test.

Another "cusum" type statistics has been introduced by B. CHALMOND [10] for solving the same problem as I.V. NIKIFOROV [42]: jump in the mean value for an ARMA (p,q) process, all the parameters of which are unknown; but in a nonparametric framework, in the sense that the knowledge of the parameters of the model is not required (see also section IV.1/*b)*). This statistics is an

extension of that of R.L. BROWN, J. DURBIN and J.M. EVANS [7] which is a cumulative sum of the deviations with respect to the model: "constant mean in the independent case": if $(Y_n)_n$ is the observed signal, the considered statistics is the following one:

$$S_n = \sum_{i=k+1}^{n} (Y_i - \overline{Y}_{i-1}),$$

where $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and where k will be precised later on; the test consists in setting the alarm as soon as:

$$\max_{k \leq n \leq N} \frac{|S_n|}{\hat{\sigma}\sqrt{N-k}} \geq \lambda,$$

where $\hat{\sigma}$ is an estimate of σ, σ^2 being the generalized variance of the deviation process $d_i = Y_i - \overline{Y}_{i-1}$:

$$\sigma^2 = E(d_n^2) + 2 \sum_{h=1}^{+\infty} E(d_n\, d_{n+h});$$

here n is greater than k which is an integer such that the process $(d_i)_{i>k}$ can be assumed to be stationary. B. CHALMOND empirically studied the probability of false alarm and the power of this test, before applying it to real data (evolution of the amount of ozone in Los Angeles).

c) Kalman filter - based approaches.

• N. ISHII, A. IWATA and N. SUZUMURA [30] investigated the segmentation of EEG signals with the aid of a Kalman filter operating on an ARMA model, only the autoregressive coefficients of which are time-varying. This algorithm was introduced by T. BOHLIN [4].

Let :

$$\begin{cases} y(t) = x(t) + c_1\, x(t-1) + \ldots + c_n\, x(t-n) \\ x(t) + a_1(t)\, x(t-1) + \ldots + a_n(t)\, x(t-n) = \lambda\, e_o(t) + K(t) \\ a_i(t) = a_i(t-1) + q_i\, e_i(t) \qquad (1 \leq i \leq n) \\ K(t) = K(t-1) + q_{n+1}\, e_{n+1}(t) \end{cases}$$

be the considered model, where the $e_i(t)$ are independent white noises.

For given n, m, $c_1,\ldots,c_m$, the parameters λ, $q_1,\ldots,q_{n+1}$ are estimated by maximum likelihood, and then $a_1(t),\ldots,a_n(t)$ and K(t) are recursively estimated by Kalman filtering. Abrupt changes are detected when the square of the (normalized) state estimation error exceeds a threshold, or when the likelihood on a small observation window is less than a threshold. The detections of the trends, or slow variations, involve the sum of the squares of the autocorrelation coefficients of the observations.

• B. WITTENMARK [59] proposed an estimate of variable parameters working at two levels: a coarse estimator, based upon a list of a priori fixed models and a decision function allowing the choice of the best fitting model (for example, the model which maximises the a posteriori distribution of the parameters); and a finer estimator based upon the Kalman filter corresponding to the chosen model. The main drawback of this method lays in the choice of the list of models and in the difficulty in discriminating these models. However, this method seems to be useful in adaptive and dual control, and B. WITTENMARK studied two examples of failures for ARMA (1,1) processes coming from the stochastic adaptive control literature.

M. MILLNERT [39b] used a similar approach to study the problem of recursive identification of abruptly changing systems. Considering an ARMA (p,p) process, the coefficients of which are allowed to take one of m possible values in $\mathbb{R}^{2p}$, he introduced two identification procedures for such systems. More precisely, if:

$$y_t = \theta'_t \phi_t + e_t,$$

where: $\theta'_t = [a_1(t),\ldots,a_n(t),\ b_1(t),\ldots,b_n(t)]$

and: $\phi'_t = [-y_{t-1},\ldots,-y_{t-n},\ u_{t-1},\ldots,u_{t-n}]$,

the identification procedures of the parameter vector are of the form :

$$\hat{\theta}^{(k)}_{t+1} = \hat{\theta}^{(k)}_t + \gamma_t\ \varepsilon^{(k)}_t\ \phi_t\ \mu^{(k)}_t,$$

where $\varepsilon^{(k)}_t = y_t - \hat{\theta}^{(k)}_t\ \phi_t$ is the prediction error according to model k, and where $\mu^{(k)}_t$ is chosen as below:

- either use the minimal prediction error, i.e. take:

$$\mu_t^{(\ell)} = \begin{cases} 0 & \text{for } \ell \neq k \\ 1 & \text{for } \ell = k \end{cases} \qquad \text{when:}$$

$$(\forall\ \ell \neq k) \quad \left[\varepsilon_t^{(k)}\right]^2 < \left[\varepsilon_t^{(\ell)}\right]^2 ;$$

- or use the a priori probabilities of the different possible states of the parameter vector θ. The computations so required can be thought of as a non linear filtering of the residuals $\varepsilon_t^{(k)}$.

The convergence properties of the two algorithms are investigated, and numerical simulations are presented; but no real example is given.

d) A problem of control.

To conclude this section devoted to failures in ARMA models, let us examine a bayesian approach for stochastic control in this framework. D.D. SWORDER and V.G. ROBINSON [51] built up regulators adapted to vectorial systems submitted to jumps in parameters. Let the system to be controlled be modelled by the linear differential equation:

$$\begin{cases} \dot{y}_t = A(t)\, y_t + B(t)\, v_t & (0 \leq t \leq T) \\ y_o \text{ given} \end{cases}$$

and let us suppose that the matrices A(t) and B(t) depend upon a markovian jump process r(t), for which a speed measure is given:

$$\begin{cases} \dot{\rho}_t = f(y_t, v_t, r(\rho_t)) & (0 \leq t \leq T) \\ \rho_o = 0 \end{cases}$$

where f is upper and lower bounded and continuously differentiable with respect to (y_t, v_t). Let be $z_t = (t, y_t, r(\rho_t))$ and let $v_t = \bar{u}(z_t)$ be the used feedback control. In the case where f weakly depends upon x_t and v_t, D.D. SWORDER and V.G. ROBINSON derived the optimal control, which, for each z_{t_o}, minimises the fonctional:

$$J(\bar{u}, z_{t_o}) = E\left[\int_{t_o}^{T} L(\tau,\ y(\tau; t_o, x_o, \bar{u}),\ \bar{u}(z_\tau)\, d\tau \mid z_{t_o}\right]$$

where L is continuous and continuously differentiable with respect to x and v, and where $y(t;\ t_o, x_o, \bar{u})$ is the root of:

$$\begin{cases} \dot{y}_t = A(r(\rho_t))\ y_t + B(r(\rho_t))\ \bar{u}(z_t) \\ y_{t_o} = y_o \end{cases}$$

for the given initial condition $(t_o, y_o, r(t_o))$. In fact, for the chosen special case, the computing algorithm of the optimal control only requires the resolution of ordinary differential equations (and not partial derivatives equations as in the general case) and is a first step for computing the optimal control in the general case.

V. ANALYSIS OF FAILURES IN LINEAR SYSTEMS MODELLED WITH STATE VARIABLES.

A very detailed bibliographical study of this problem has already been presented in [54] by A.S. WILLSKY who, for the first time as it seems, precisely formulated the failure detection problem in this framework, and presented a large set of both practical and theoretical concepts concerning the ways of approaching it, as well as some heuristical comparisons between the various mentioned methods, especially from the point of view of robustness and tradeoff in complexity versus performance, as it has been presented in section I.

This paragraph is mainly devoted to complete this study by more recent algorithms; but, in order to preserve the coherence of the presentation, some of the methods mentioned by A.S. WILLSKY will be recalled here.

The adopted order of presentation is the following one: methods based upon analysis of innovations of Kalman filters; multiple models methods; use of nonlinear filtering; use of redundancy in information.

But for a few exceptions which will be explicitly mentioned, the studied systems are here vectorial ones.

1/ Innovations-based algorithms.

As we already mentioned in section IV, failure detection can be achieved by more or less sophisticated processing of innovations of Kalman filters. Detection of jumps in the mean of the signal can be achieved by looking for jumps in mean of the innovations. For failures in the dynamics of the system, jumps in the square of the innovations are of interest; but it should be noticed that this kind of extension is non trivial because the dynamics of a Kalman filter are difficult to be identified. Finally, when the effect of the failure on the observations is known and can be modelled, a more accurate detection can be obtained by using sequential probability ratio tests or generalized likelihood ratio tests.

a) Tests for centering or whiteness of the innovations.

Ten years ago, R.K. MEHRA and J. PESCHON [39] introduced some elementary failure detectors based upon the innovations of the Kalman filter adapted to the model which describes the system in its normal operating mode. These tests are based upon the fact that any type of failure (measurement bias, measurement noise, change in the dynamics of the system, change in the level of the input noise, even change in the system structure) makes the normalized innovations loose their fondamental properties: zero mean, variance 1, whiteness.

Tests for whiteness classically use the autocorrelation function $(c_k)_{k\geq 0}$ and its asymptotical property of normality.

The centered feature of the innovations (γ_n) may be tested either with the aid of the asymptotical normality of their empirical mean: $\bar{\gamma} = \frac{1}{N} \sum_{n=1}^{N} \gamma_n$, or with the quadratic form related to:

$$\hat{c}_o = \frac{1}{N} \sum_{n=1}^{N} (\gamma_n - \bar{\gamma})(\gamma_n - \bar{\gamma})' \quad \text{and to } \bar{\gamma}:$$

$$T^2 = N\, \bar{\gamma}'\, \hat{c}_o^{-1}\, \bar{\gamma}.$$

Finally, the unity variance of the normalized innovations may be tested with the asymptotic distribution of $\hat{c}_o$ which is Wishart's one, or of the trace of $\hat{c}_o$ which is a χ^2 with $(N-1)p$ degrees of freedom -(p = dimension of the innovations)-.

The reader will find other examples of tests for whiteness, parametric or not, in time or frequency domains, in [35a] (chap. 4).

b) Sequential probability ratio tests (SPRT).

Some examples of use of such tests, in order to corroborate elementary failure detections, will be found in paragraph V.4/. Let us examine here some algorithms in which the SPRT are involved at a different level.

First, a rather old example proposed by P.M. NEWBOLD and Y.C. HO [41] who studied a simplified version of a SPRT of the hypothesis H_o: "*a failure*" against the hypothesis H_1: "*no failure is present*", for the case of a change in

the variance of the input noise for scalar dependent observations, and applied it to a gyro inertial system. This simplification tends to reduce the amount of computations which, for such a test, goes increasing with the number of observations. The proposed test is then, in fact, a SPRT of the hypothesis H_o: "*no failure, i.e. all the observations are governed by the law* P_o" against the hypothesis H'_1: "*all the observations are governed by the law* P_1", and which furthermore takes into account the informations concerning the system which are known before the activation of the test. The experimental results showed that the decrease in performance, which this involves, is acceptable, at least for some applications.

More recently, T.T. CHIEN and M.B. ADAMS [12] have designed a system for detection and isolation of failures in order to test the reliability of the inertial measures of a space shuttle. This detector, based upon a SPRT of H_0 against H'_1 with "feedback", has been obtained by minimizing the mean time delay for detection of a failure under some constraints concerning the probabilities of false alarm and nondetection. In fact, the idea of this "feed-back" has already been introduced by A.N. SHIRYAEV (see [2], [3b], and section II.2/*b)* and *figure n°2*): in order not to degrade the detection of a failure occuring after a long observation time, the threshold, that the SPRT has to cross over, is updated from time to time according to the behaviour of the test statistics.

Finally, T. YOSHIMURA et al. [61] introduced a detector which classically connects a filter corresponding to the normal operating mode of the observed system and a failure detector based upon a SPRT of H_0 against H'_1 for each component of the innovation, to which they joined an extended adaptive Kalman filter designed to estimate the unknown parameters and the states when the failure hypothesis has been accepted. The only example of application they considered was synthetical and concerned with a scalar observation, with three dimensional state space, and state noise only on the first component.

c) Generalized likelihood ration tests (GLR).

This test is a sequential probability ratio test of the hypothesis H_0: "*no failure, i.e. all the observations until time* n *are governed by the law* P_o",

against the hypothesis H_1: *"a failure is present, i.e. the observations are governed by the law* P_0 *until time* r, *and by the law* P_1 *from time* (r+1) *to time* n". But, the time r and the magnitude of the failure being unknown, they are replaced, in the likelihood under H_1, by their maximum likelihood estimates.

This test, which is a classical one in Statistics, has been introduced in failure detection first in the aerospatial domain ([45], [55]) and first presented in its more general form by A.S. WILLSKY and H.L. JONES ([56], [57]) whose presentation will be kept. A number of applications of this algorithm, possibly including some modifications, can be mentioned in domains as various as aerospatial ([8], [11a], [11b], [13]), electrocardiography ([25a], [25b]), urban traffic supervision [58], geophysics [3c].

Consider a system which, under a possibly disturbed operating mode, can be modelled by the following equations:

$$\begin{cases} x_{k+1} = \phi(k+1,k)\, x_k + \Gamma_k\, w_k + \delta_{r+1,k+1}\, \nu \\ y_{k+1} = H_{k+1}\, x_{k+1} + v_{k+1} \end{cases}$$

where the state $x_k \in \mathbb{R}^n$, the observation $y_k \in \mathbb{R}^p$, and $\{w_k\}$ and $\{v_k\}$ are two gaussian, zero mean, white and independent sequences, with respective covariance matrices Q_k et R_k (positive). ν is the magnitude of the failure.

The GLR algorithm is essentially devoted to the detection of failures which have a "linear" effect on the system, of the type jump or ramp on the state, jump or ramp on one or more captors, as is shown on the previous model, and seems to be of poor interest for detecting failures in the dynamics or changes in the parameters of the system (in ϕ, Γ, H), because introducing these failures in an augmented state espace for example leads to the loss of the linearity of the state model, which is precisely the key point in the elaboration of the algorithm (additive decomposition of the innovation). This method is interesting because the estimates of the failure time and of the magnitude of the failure and the generalized likelihood ratio can be computed in a recursive manner, as it will be shown later; furthermore it provides a compensation scheme for updating the filter estimates after a system change has been detected.

Under the null hypothesis H_0 ($r=+\infty$), the Kalman filter is described by the following equations:

$$\begin{cases} \hat{x}_{k+1|k} = \phi(k+1,k)\ \hat{x}_{k|k} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\ \gamma_k \end{cases}$$

where γ_k is the innovation:

$$\gamma_k = y_k - H_k\ \hat{x}_{k|k-1},$$

and where the gain, the error covariance and the covariance of the innovation are given by:

$$\begin{cases} K_k = P_{k|k-1}\ H'_k\ V_k^{-1} \\ P_{k+1|k} = \phi(k+1,k)\ P_{k|k}\ \phi'(k+1,k) + \Gamma_k\ Q_k\ \Gamma'_k \\ P_{k|k} = (I - K_k\ H_k)\ P_{k|k-1} \\ V_k = H_k\ P_{k|k-1}\ H'_k + R_k. \end{cases}$$

Now, x_k, $\hat{x}_{k|k}$, y_k and γ_k can be expressed in a form that explicitly involves r and ν. Indeed, because of the linear effect of the failure, we can write:

$$\begin{cases} x_k = x_k^1 + x_k^2 \\ y_k = y_k^1 + y_k^2 \end{cases} \qquad \begin{cases} \hat{x}_{k\,k} = \hat{x}_{k\,k}^1 + \hat{x}_{k\,k}^2 \\ \gamma_k = \gamma_k^1 + \gamma_k^2 \end{cases}$$

where the upperscript 1 indicates all the effects but r and ν, and the upperscript 2 indicates <u>only</u> the r and ν effects. From the filter equations, we deduce:

$$\begin{cases} x_k^2 = \phi(k,r+1)\ \nu \\ y_k^2 = H_k\ \phi(k,r+1)\ \nu \\ x_{k|k}^2 = F(k,r+1)\ \nu \\ \gamma_k^2 = G(k,r+1)\ \nu \end{cases}$$

where:
$$\begin{cases} F(k,r) = K_k\ G(k,r) + \phi(k,k-1)\ F(k-1,r) \\ G(k,r) = H_k[\ \phi(k,r) - \phi(k,k-1)\ F(k-1,r)] \\ G(k,k) = H_k \end{cases}$$

(For $k \leq r$, ϕ, F, G are obviously zero).

G is called the signature of the failure.

Therefore, the two hypothesis to be tested are:

H_0: $\gamma_k = \gamma_k^1$

H_1: $\gamma_k = \gamma_k^1 + G(k,r+1)\ \nu$

where γ_k^1 is a white gaussian sequence, with covariance matrix V_k. The generalized likelihood ratio test computes the likelihood ratio:

$$\Lambda_k = \frac{L(\gamma_1,\ldots,\gamma_k \mid H_1,\ r=\hat{r}_k,\ \nu=\hat{\nu}_k)}{L(\gamma_1,\ldots,\gamma_k \mid H_0)}$$

where $\hat{r}_k$ and $\hat{\nu}_k$ are the maximum likelihood estimates of r and ν under the hypothesis H_1:

$$(\hat{r}_k,\hat{\nu}_k) = \arg\max_{(\tilde{r},\tilde{\nu})} L(\gamma_1,\ldots,\gamma_k \mid H_1,\ r=\tilde{r},\ \nu=\tilde{\nu}).$$

Under gaussian assumptions, the log-likelihood ratio is:

$$\ell(k,\hat{r}_k+1) = 2\ \mathrm{Log}\ \Lambda_k$$

$$= \sum_{j=1}^{k} \gamma_j'\ V_j^{-1}\ \gamma_j - \sum_{j=1}^{k} (\gamma_j - G(j,\hat{r}_k+1)\hat{\nu}_k)'\ V_j^{-1}\ (\gamma_j - G(j,\hat{r}_k+1)\hat{\nu}_k)$$

and $\hat{\nu}_k$ minimizes:

$$f(y) = \sum_{j=1}^{k} (\gamma_j - G(j,\hat{r}_k+1)y)'\ V_j^{-1}\ (\gamma_j - G(j,\hat{r}_k+1)\hat{\nu}_k).$$

So: $$\hat{\nu}_k = C^{-1}(k,\hat{r}_k+1)\ d(k,\hat{r}_k+1)$$

where: $$C(k,\theta) = \sum_{j=\theta}^{k} G'(j,\theta)\ V_j^{-1}\ G(j,\theta)$$

and: $$d(k,\theta) = \sum_{j=\theta}^{k} G'(j,\theta)\ V_j^{-1}\ \gamma_j.$$

Besides $\hat{r}_k$ is the time between 0 and (k-1) which maximizes the log-likelihood ratio:

$$\ell(k,\tilde{\theta}) = d'(k,\tilde{\theta})\ C^{-1}(k,\tilde{\theta})\ d(k,\tilde{\theta}).$$

The decision rule is as follows:

$$\ell(k, \hat{r}_k+1) \underset{H_0}{\overset{H_1}{\gtrless}} \varepsilon.$$

The first drawback of this algorithm is the increasing length of the bank of filters it requires (searching $\hat{r}_k$ between 0 and k-1). In practice, the search is restricted to a finite window of length M:

$$k-M \leq r_k \leq k-1.$$

(In fact, $k-M \leq \hat{r}_k \leq k-n$, where n is the state space dimension). The choice of the length M of the window and of the threshold ε is a difficult task. We have shown in [3c] that, in some cases, ε should increase approximately linearly with M, and that the number of false alarms and nondetections may be highly depending upon ε for noisy signals. So a modified version of the algorithm has been designed (see [3c]). It involves a minimal magnitude of failure to be detected, and seems to be less sensitive to the choice of the different parameters of the detector, which allow, furthermore, an easier adjustment by the practical user.

Let us also notice that a simplified version of the GLR algorithm has been investigated by E.Y. CHOW [13]: when the failure time is not of main interest, the maximization on r is cancelled and the estimate $\hat{r}_k$ is constrained to the value k-M .

2/ Multiple models methods.

This very fruitful approach uses a bank of filters, each of which corresponds to a given direction of failure, and the decision function is based upon the a posteriori probabilities of the different failures conditionally to the observations. These probabilities can be sequentially computed as functions of the innovations of the different filters (see [55]).

However, apart from the fact that this method requires a good knowledge of the list of the possible failures, its drawback is to involve a bank of filters, the size of which is exponentially increasing with time: indeed

the a posteriori law of the state governed by the equation:

$$x_{k+1} = \phi(k+1,k)\ x_k + w_k$$

where the variance of the noise w_k takes one of s values with probabilities p_i $(1 \leq i \leq s)$, is given by:

$$p(x;k) = \sum_{i_o=1}^{s} \cdots \sum_{i_{k-1}=1}^{s} p^i\ N(x;\ \eta_i, P_i),$$

where p^i is the conditional probability that w_{i_j} has the state i_j $(1 \leq j \leq k-1)$ and η_i, P_i are the mean and the variance which are estimated by the Kalman filter corresponding to the noises w numbers i_j $(1 \leq j \leq k-1)$.

In practice, the use of heuristical methods allows to suppress the exploration of many branches of the tree. An approximated method, where the failure hypothesis is only introduced every n steps, will be found in [55].

Recall that two algorithms based upon multiple hypothesis have already been mentioned in section IV.2/*c)*. (See [59], [39b]).

Furthermore, J.K. TUGNAIT and A.H. HADDAD [53] introduced probabilities of transition between the various possible states of the system, which are supposed to be in a finite number. They investigated the estimation of the state of the system:

$$\begin{cases} x_{k+1} = A\ x_k + B\ w_k \\ y_k = C\ x_k + v_k \end{cases},$$

with the criterion of minimum mean square error, when the state noises w_k and the measurement noises v_k are governed by one of the S gaussian multivariate distributions of an a priori given set which is provided with a S XS transition matrix π. In other words, the considered estimate has the form:

$$\hat{x}_{k|k} = \sum_{I_j(k) \in \Omega_k} \hat{x}^{(j)}_{k|k}\ P(I_j(k)\,|\,Y_k),$$

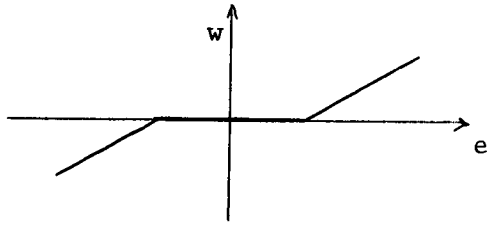
where $I_j(k)$ is a sequence of $\{i_o,\ldots,i_k\}$ of index of the noise distributions, and: $\hat{x}^{(j)}_{k|k} = E(x_k\,|\,Y_k, I_j(k))$. J.K. TUGNAIT and A.H. HADDAD then tried to

limitate the exponential growing of the size of the memory, and proposed an approximated estimate which reduces the number of computations by the choice, at each time, of the number of sequences of possible states for the system: only the sequences, for which the normalized prediction error is less than a threshold, are kept. The considered example of application is synthetical, the state and the observation are scalar, and the system has only two possible states.

Finally, B. FRIEDLAND [20] studied the system:

$$\begin{cases} x_{k+1} = x_k + w_{k+1} \\ y_k = H_k x_k + v_k \end{cases}$$

where $\{v_k\}$ is a white sequence of gaussian variables with zero mean and covariance matrix R, and $\{w_k\}$ is a white sequence of random variables, the probability law q of which has a Dirac mass at the origine (for example, w_k, which models the jumps on the state, is the output of a nonlinear filter which is excited by a white noise e -see the figure-).



B. FRIEDLAND derived the maximum likelihood estimate for the state (conditionally to the observations) and proposed a recursive approximated solution for the two boundary-values problem which is involved. Indeed the estimate: $X_n = (x_1^n,\ldots,x_k^n,\ldots,x_n^n)$, which maximizes the a posteriori law $p_n(x_1,\ldots,x_n|Y_n)$, is given by the following system of equations:

$$\begin{cases} \gamma(x_k^n - x_{k-1}^n) = \lambda_{k-1}^n & (1 \leq k \leq n-1) \\ \lambda_{k-1}^n - \lambda_k^n = H'_k R^{-1}(y_k - H_k x_k^n) \end{cases}$$

where $\gamma(\omega) = (\gamma_1(\omega),\ldots,\gamma_n(\omega))'$ and $\gamma'_i(\omega) = \left.\frac{\partial[\text{Log } q(\omega)]}{\partial\omega}\right|_{\omega_i}$, with the boundary conditions:

$$\begin{cases} \lambda_o^n = \dfrac{\partial\,[\mathrm{Log}\ p_o(x_o)]}{\partial x_o} \\ \lambda_{n-1}^n = R^{-1}(y_n - H_n\, x_n^n), \end{cases}$$

$p_o(x_o)$ being the a priori density of the initial state x_o. The approximated solution was derived by an analogy with a problem of an electrical network, and its validity tested via a simulation study.

3/ Use of nonlinear filtering.

• In order to study failures in a linear scalar system of the form:

$$\begin{cases} x_{k+1} = A_i\, x_k + B_i\, u_k + w_k(i) \\ y_k = C_i\, x_k + v_k(i) \end{cases}$$

which is characterized by two sets of parameters (i=0,1), K. GIRIDHARAGOPAL et al. [23] introduced a nonlinear stochastic model:

$$\begin{cases} x_{k+1} = [A_0 + g_{k+1}(A_1 - A_0)]\,x_k + [B_0 + g_{k+1}(B_1 - B_0)]\,u_k + w_k(g_{k+1}) \\ g_{k+1} = f(g_k, x_k) + \psi_k \\ y_k = [C_0 + g_k(C_1 - C_0)]\,x_k + v_k(g_k) \end{cases}$$

where f is nonlinear, and they used an extended Kalman filter to estimate the state variable x and the state (0 or 1) of the system; i.e. to compute $\hat{x}_{k|k}$ and $\hat{g}_{k|k}$. This filter has been compared to other filters of the same type via a simulation study.

• On the other hand, M.H.A. DAVIS [15] applied nonlinear filtering techniques to a problem of failure in the state dynamics of a linear scalar system described by the following stochastic differential equations:

$$\begin{cases} dx_t = a(t)\, x_t\, dt + g(t)\, dw_t \\ dy_t = h(t)\, x_t\, dt + dv_t \end{cases},$$

where: $a(t) = a_0(t)\,(1-z_t) + a_1(t)\, z_t$, and $z_t = 1_{\{t \geq T\}}$.

He derived an implementable approximation of the optimal (infinite dimensional) filtering equations, which allows computing the estimates of the state x_t and of the failure time (by estimating z_t). In practice, it seems that, as for A.N. SHIRYAEV's optimal algorithm (see [2]), the choice of the threshold, to which the a posteriori probability is compared, is a difficult task.

4/ Use of redundancy in information.

This approach is mainly encountered in the aerospatial domain. Only three examples will be mentioned in this paragraph ([13], [16], [34]).

M. LABARRERE et al. [34] designed a failure detector for a double set of captors in a plane, which explicitly used redundancy relationships. These relations, which do not depend upon atmospherical perturbations, allow computing each measured quantity as a function of the others and their derivatives, and estimating the failure time by a logical test on the difference between the two measures of a same quantity. This elementary detector is strenghtened by three sequential probability ratio tests (see section V.1/*b)*), the first one devoted to corroborate the presence of a failure and the other two to isolate the failing captor.

J. DECKERT, A.S. WILLSKY et al. [16] used a similar approach to detect and identify captor failures in a doubly redundant environment; the maximal exploitation of the functional relationships between the various captors is connected to several sequential probability ratio tests. The direct comparison between the two captors of a same pair allows a first failure detection and the activation of several SPRT which involves this direct redundancy in the pair as well as all the possible analytic redundancy between the type of the failing instrument and all the other types of instruments. The main advantage of such an approach is the noticeable reduction of the mean computing burden.

The difference between these two approaches lays in the level at which the redundancy between all the types of instruments is involved: M. LABARRERE [34] used it from the estimation of the measured quantities,

and thus before the elementary test on the difference between the two elements of a pair of captors; whereas J.C. DECKERT et al. [16] only introduced in the SPRT used to corroborate a failure and refine its identification, the elementary test being only preceded by a moving window filtering for each type of measures.

Recently, E.Y. CHOW [13] presented a general formulation of the concept of analytical redundancy in terms of parity relations, which are used to design a robust residual-generation process (taking into account noises and modelling errors). These residuals are then analyzed with the aid of various decision rules, which have been derived from a sequential bayesian point of view and the performances (risks and detection probabilities) of which have been numerically studied.

VI. SOME APPROACHES FOR CHANGES IN MORE GENERAL MODELS.

1/ Nonlinear systems.

The study of nonlinear systems submitted to changes in their structure has been approached by V.A. BUKHALEV [9], who considers scalar systems described by nonlinear differential equations corresponding to various states of the model, where the different states are supposed to form a Markov chain. He investigates the estimation of the state variable of the system and of the state of the model by mean of the least squares method and the a posteriori mode.

Let us also recall the work of A.S. WILLSKY and others [58], which was mentioned in the last section, and in which failures in nonlinear systems are analyzed with the aid of a linearized Kalman filter connected to a generalized likelihood ratio detector or a multiple model detector. We shall come back to that study in section VII for a comparison.

2/ Changes when the observations are scalar, independent, and governed by an arbitrary law.

• When the probability laws before and after the failure are known, and the later is absolutely continuous with respect to the former, T. BOJDECKI [5] derived the stopping time which is optimal in the sense that it maximizes the probability that the "delay" for detection is, in absolute value, less than a given threshold. He worked with the little constraining further hypothesis that the failure time is governed by an a priori geometrical law.

• Still in a sequential framework and when the probability laws before and after the failure are supposed to be known, S. ZACKS [62] numerically determined the distributions of two stopping-times used as failure detectors: one of them, which has been introduced by A.N. SHIRYAEV in a bayesian framework (see [2]), is the first time at which the a posteriori probability, that

the failure time occurs before it exceeds a threshold; the other one is the first crossing-time of a linear level by a cumulative sum of the observations. The validity of the approximations used for the numerical evaluation of these distributions has been tested for a change in the parameter of a Poisson's law, as was mentioned in section II.

• From a nonsequential point of view, B.S. DARKHOVSKII [14a] studied a nonparametric method for a posteriori detection of a change between two laws, the distribution functions of which (F_0 and F_1) are continuous and such that:

$$(\alpha) \quad \int_{-\infty}^{+\infty} F_1(y)\, dF_0(y) \neq \frac{1}{2}.$$

This method, based upon Mann-Whitney's statistics, is consistent, i.e., when the number of observations tends to infinity, the estimate converges to the true failure time. This estimate $\hat{r}$ is the time r which maximizes (or minimizes, according to the "direction" of the change) the Mann-Whitney's statistics:

$$V_N(r) = \frac{1}{r(N-r)} \sum_{k=r+1}^{N} \sum_{i=1}^{r} 1_{\{Y_i > Y_k\}}.$$

An analogous test has been studied for discrete observations taking a finite number K of values ([14b]). It seems that, in the case of continuous laws, using a sampling on a finite number of levels and the "discrete" test gives better results than the "continuous" test. One reason for this is the fact that two continuous distributions are not always distinguishable in the Mann-Whitney's sense (i.e. (α)), when discrete distributions are always distinguishable with respect to the criterion:

$$\sum_{i=1}^{K} (p_i - q_i)^2 > 0.$$

It has to be noticed that, contrary to T. BOJDECKI's stopping-time, this statistics does not require the knowledge of the probability laws before and after the failure.

3/ An example of change when the observations are vectorial, dependent, and governed by an arbitrary law.

Let us first specify that J. SEGEN and A.C. SANDERSON's approach [46], which is presented here, allows to rediscover, in the case of a scalar auto-regressive process, the test statistics proposed by L.I. BORODKIN and V.V. MOTTL' [6] which has been introduced in section IV.1/ . This relatively general new test is of "cusum" type; it involves cumulative sums of entropy and allows the detection of any change which does not decrease the prediction entropy. More precisely, let $Y_1,\ldots,Y_n,\ldots$ be a sequence of observations described by the predictive probabilities $g_o(Y_n|Y^{n-1})$,

where: $Y^{n-1} = (Y_{n-1},Y_{n-2},\ldots,Y_1)$, and let be: $X_n = -\text{Log } g_o(Y_n|Y^{n-1})$.

The considered statistics U_n is the cumulative sum of the innovations of this sequence $(X_n)_n$; in other words:

$$U_n = \sum_{i=1}^{n} T_i$$

where: $T_i = \int g_o(x|X^{n-1}) \text{ Log } g_o(x|X^{n-1})dx - \text{Log } g_o(X_n|X^{n-1})$.

J. SEGEN and A.C. SANDERSON showed that, under H_0, $\{U_n\}_n$ is a martingale sequence; under H_1, and with the hypothesis that the new law g_1 after the change verifies:

$$\int g_1(x|X^{n-1}) \text{ Log } g_1(x|X^{n-1})dx \leq \int g_o(x|X^{n-1}) \text{ Log } g_o(x|X^{n-1})dx,$$

U_n can be uniquely decomposed as:

$$U_n = M_n + Q_n,$$

where $\{M_n\}_n$ is a martingale and $\{Q_n\}_n$ a predictable sequence (i.e. Q_n is a function of X^{n-1}) such that $Q_n=0$ for $n\leq r$ and $Q_n>Q_{n-1}$ for $n>r$. It follows that, under H_0 and if $\{T_n\}$ is furthermore stationary and ergodic and if:

$$Z_n = \frac{1}{\sigma} U_n, \qquad \sigma^2 = \text{var } (T_n),$$

and: $$W_n(t) = Z_{[nt]},$$

then $\frac{1}{\sqrt{n}} W_n(t)$ converges in distribution towards the standard brownian motion.

(N.B.: if $\{T_n\}$ is nonstationary, consider:

$$Z_n = \sum_{i=1}^{n} \frac{T_i}{\int g_o(x|X^{i-1}) \operatorname{Log}^2 g_o(x|X^{i-1})dx - \left[\int g_o(x|X^{i-1}) \operatorname{Log} g_o(x|X^{i-1})dx\right]^2}).$$

Furthermore, under H_1 the mean of U_n increases. The test for failure thus consists in comparing Z_n to a threshold of the form $\sqrt{m}\, g(\frac{n}{m})$, where m is the supposed order of magnitude of r (in other words, we come back to the test problem:

$$H_0^{(m)}: r>m \qquad \text{against} \qquad H_1^{(m)}: r\leq m),$$

and where g has to be a not too rapidly increasing function in order to ensure a quick detection and a test power less than unity (for example: $g(t) = \sqrt{2t \operatorname{Log} \operatorname{Log} t}$). For an autoregressive process:

$$Z_n = \frac{1}{\sqrt{2}} \sum_{i=1}^{n} (\frac{e_i^2}{\sigma^2} - 1),$$

where e_i is the one-step prediction error (see [6] and section IV.1/).

This test for failure has been used by A.C. SANDERSON and J. SEGEN [44] for the segmentation of electroencephalograms.

VII. SOME COMPARATIVE STUDIES.

All the authors do not bother comparing the algorithm they propose to competitive detectors. Therefore, the results of the comparative studies which will be presented in this section are concerned with a few algorithms from the list which has been studied in the previous ones. The preceding order of presentation will be kept here.

1/ Changes in mean.

• In the case of a change in the parameter of a Bernoulli distribution, A.N. PETTITT [43] compared two test statistics used in a nonsequential framework: one base upon cumulative sums of the observations, the other built from the likelihood ratio. In section II.2/*a)*, we already mentioned that these two tests have nearly the same power.

On the other hand, A.N. PETTITT also compared, from an asymptotic point of view as well as for small samples, two estimates of the failure time: the maximum likelihood one, and another one based upon the previous "cusum" type statistics (see section II.2/*a)*). In the "symmetric" case (the two Bernoulli parameters are symmetric with respect to $\frac{1}{2}$), the asymptotic distributions of these two estimates are equivalent, provided that $\frac{r}{N} \to \frac{1}{2}$. The comparison on small samples did not give raise to significant differences.

• S. ZACKS's study [62], which is concerned with the case of change between any discrete known probability laws, involves a comparison of the numerical distributions of the two considered stopping-times, for the Poisson's law. (See section VI.2/). The bayesian procedure seems to tend to stop earlier the "cusum" type test.

• Finally, concerning the case of change in mean for a sequence of random independent gaussian observations, M. BASSEVILLE's comparative study [3b], via a simulation study as well as from a theoretical point of view, allowed a comparison between simple "filtered derivatives" detectors, A.N. SHIRYAEV's

simplified detector (i.e. sequential probability ratio test) and D.V. HINKLEY's detector. (See section II.2/*b)*). It has so been shown that the later is optimal, in the sense that it minimizes the mean delay time for detection of a jump (conditionally to the absence of false alarms before the jump) for a fixed mean time between false alarms. Furthermore, Hinkley's detector is more robust in so far as it is less affected than the others by a poor a priori knowledge of the signal variance and of the failure magnitude.

2/ Changes in regression models.

• P.K. SEN [47] compared the asymptotic powers of two tests based either on least squares estimates or linear rank statistics (see section III.3/). The asymptotic relative efficiency of the later with respect to the former is greater than 0.864 for any continuous probability law of the noise, reaches $\frac{3}{\pi}$ in the gaussian case, and exceeds 1 as soon as the noise law is more flat than the gaussian one.

• K. GARBADE [20] compared two tests proposed by R.L. BROWN, J. DURBIN and J.M. EVANS [7] (cumulative sums of normalized innovations or of squares of normalized innovations) to a variable regression procedure, by analyzing their powers via a simulation study. The likelihood ratio statistics seems to provide a more powerful test than the cumulative sum of normalized innovations. These two types of procedure have also been compared on economical data.

• Finally, let us mention M. BASSEVILLE and A. BENVENISTE's study [3c], who compare, on geophysical data, several failure detection algorithms, among which more elaborated "filtered derivatives" detectors, Hinkley's algorithm, Willsky's generalized likelihood ratio detector and a modified version of it, and a new algorithm mixing Hinkley's and Willsky's detectors.

3/ Changes in autoregressive models.

B. CHALMOND [10] studied jumps in mean for first order autoregressive processes, and showed that the generalized likelihood ratio test is uniformly more powerful than the test based upon the cumulative sums of the deviations with respect to a model: "constant mean in the independent case".

From a completely different point of view, S.A. LAKEHAL [35a] compared several failure tests based upon jumps in the order of the autoregressive model which is estimated with the aid of three different criteria (see section IV.1/).

4/ Changes in linear systems modelled with state variables.

It seems that the first comparative study in this domain is due to A.S. WILLSKY et al. [58] who balanced, on real data from urban traffic, the results obtained on one hand with the multiple model method, and, on the other hand, with an extended Kalman filter connected to the generalized likelihood ratio algorithm.

Recall that the comparative study [3c] described at the end of section III.2/ can also mentioned in the present one.

CONCLUSION.

In this paper, number of failure detectors for changes in various models have been presented, coming from the Automatical as well as Statistical literatures.

In practice, the choice of an adequate algorithm is strictly depending upon the considered application: the examination of the constraints it involves both in a hardware and software framework will fruitfully give a guide for solving the tradeoff of complexity versus efficiency which was mentioned in the Introduction. On the other hand, a detector can possibly be chosen for a given application only after a careful inspection of the models and the failure types which it is able to handle, of its complexity, its performance and the tradeoff it realizes between false alarms, nondetections and delay for detection of a failure (this tradeoff also depending upon the application), and finally of its robustness with respect to modelling errors. Any comparative study between different detectors should consider all these questions. But it seems that robustness is still an open problem. However, some recent studies have allowed some partial answers: among others, let us mention I.V. NIKIFOROV [42] and M. BASSEVILLE [3b] who studied the robustness of D.V. HINKLEY's algorithm (based upon the deviations, with respect to its extrema, of the cumulative sum of the observations), either in a theoretical framework for a change in mean or a change in Bernoulli parameter [3b], or by approximations and simulations for a change in mean of an ARMA process [42]. See also [3c] for another comparative study.

Let us emphasize that a theoretical study for robustness never dispenses from an experimentation of the algorithm on real data, especially if these data do not meet the assumptions made for the theoretical study; for example, D.V. HINKLEY's algorithm has been successfully used in image processing for edge detection [3a], with data for which the assumptions of independent observations and constant variance are obviously not verified; the same is true in proximetry where the used captors involve obvious nonlinearities. (see [18]).

On the other hand, robustness and complexity have also to be challenged, in order to avoid use of too expensive algorithms (for some applications) which allow a poor gain in robustness: for example, it has been recently shown [22] that nonparametric tests, of the type median, Wilcoxon's test, Haga's test, the robustness of which is well known by the statisticians, are of poor interest in image processing because of the large computing time they require and of their poor performances on highly detailed pictures.

Let us also notice that, apart from the problem of the robustness of the detectors, the choice of the models to be used for failure detection in real data is still a difficult task. However, some recent studies ([3c], [25a], [25b]) tend to indicate that using simple models disturbed by failures with linear effects (see section V.1/*c)*) is a fruitful approach.

Finally, let us emphasize that, as far as failure detection is concerned, two main types of problems can be distinguished:

- failures in signals for which either the models are well known and the main problems are quickness, accuracy, and cost of the failure detection (as in aerospatial); or the models are only a tool used for localizing the failures, and where the main problem is robustness because the involved models are often poorly validated (EEG, ECG, pictures, geophysics);

- segmentation of a signal which is approximately represented by a large amount of models, in an artificial intelligence framework (speech processing, EEG signals), in which simple models are of no use and failures are not really abrupt and therefore are recognized and accepted, using simplifying heuristics and according to a kind of dynamic programming; for this type of study, the final judgement should take into account the signal processing together with the "artificial intelligence" processing, which do not have the same nature nor the same computational weight (in speech processing for example, the former processing is about 5 % or 10 % of the whole processing, whereas the later is about 90 % or 95 %).

FIGURE CAPTION

Figure n° 1: Hinkley's algorithm.

n° 2: Shiryaev's algorithm.

n° 3: Hines' geometric moving average detector.

n° 4: A failure detector in regression models.

n° 5: A failure detector in AR models.

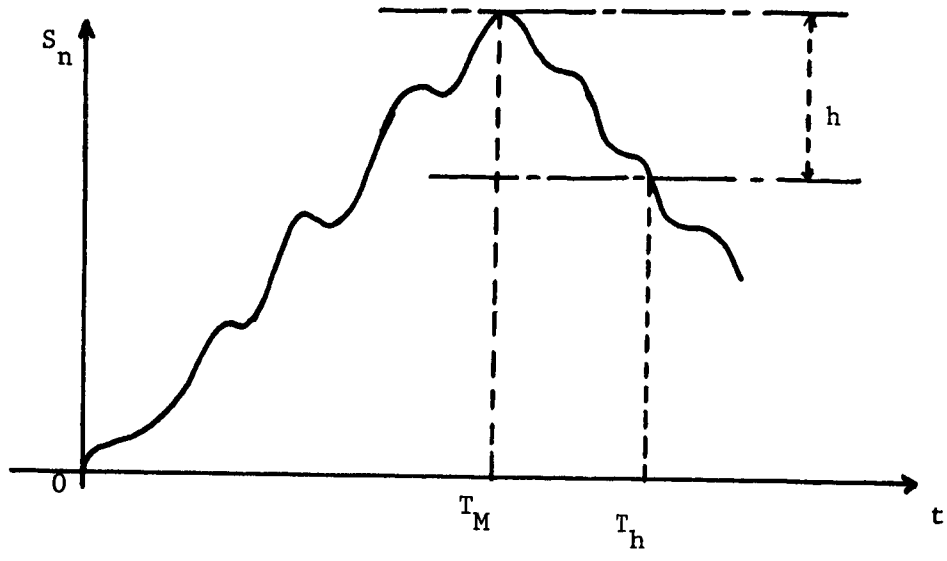


Figure n°1: Hinkley's algorithm

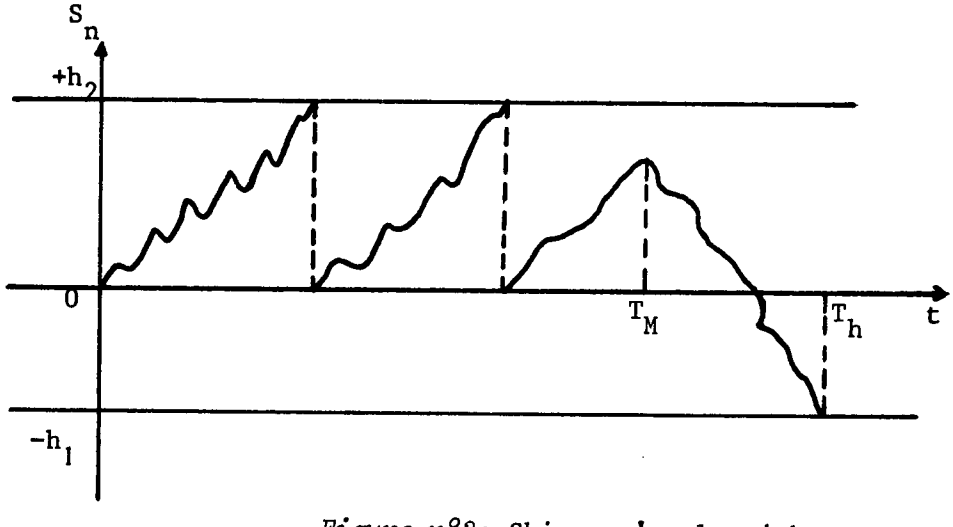


Figure n°2: Shiryaev's algorithm

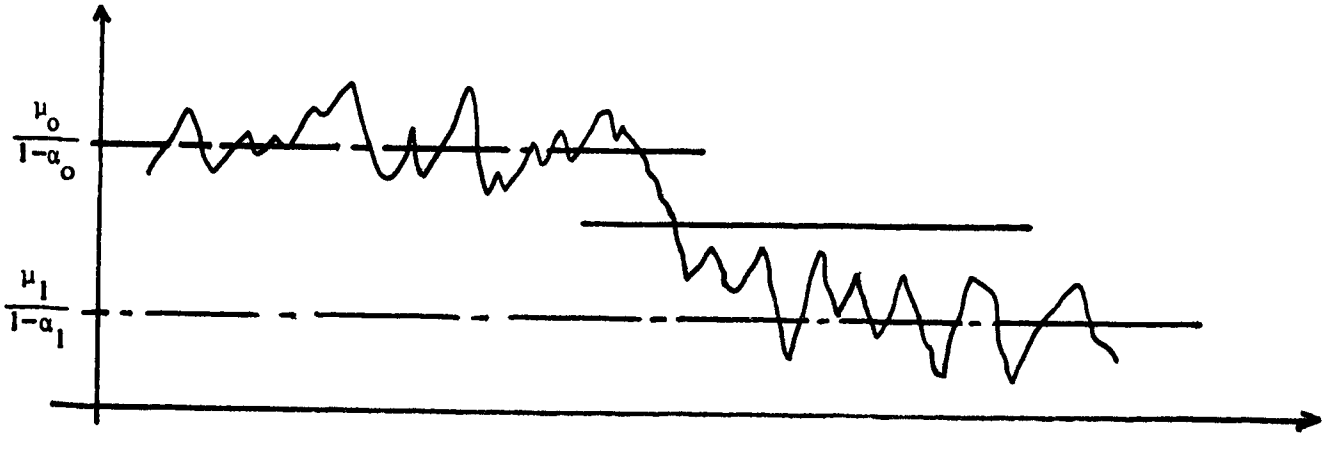


Figure n°3: Hine's geometric moving average detector

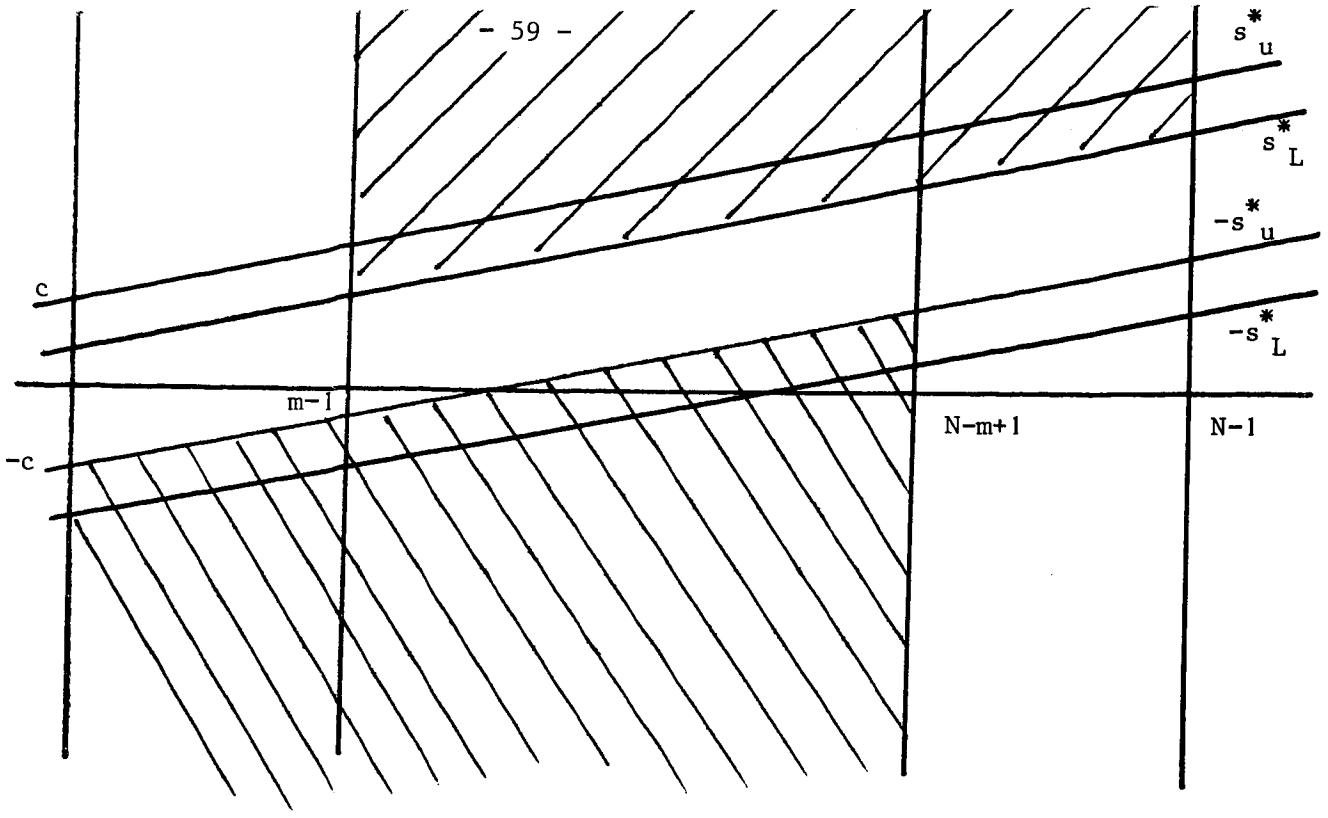


Figure n°4: A failure detector in regression models

\\\ H_0 accepted

// H_0 rejected

elsewhere: no conclusion

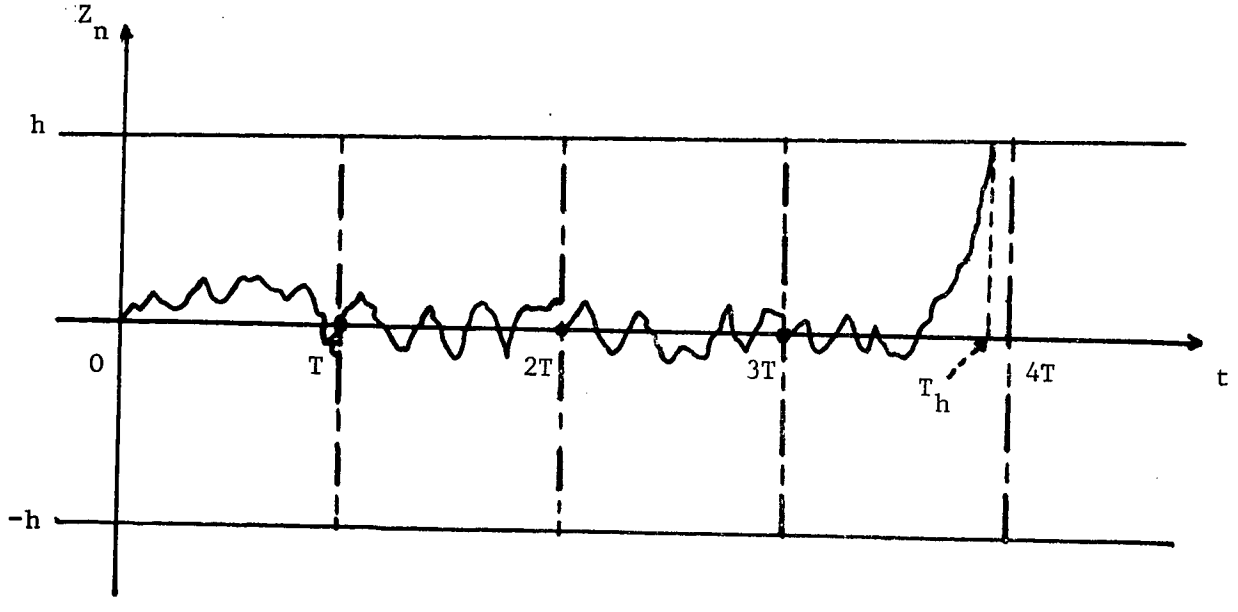


Figure n°5: No failure detector in AR models

ACKNOWLEDGEMENT

The author wishes to acknowledge helpful discussions with Pr. A.S WILLSKY (M.I.T.) and with Dr. A. BENVENISTE (I.R.I.S.A., Rennes-France) whose careful reading of an earlier version of this paper led to noticeable modifications in the structure of some sections.

BIBLIOGRAPHY

• *References mentioned in* [2], [3b], [17b], [48], [54];

[1] M. BAGSHAW, R.A. JOHNSON (1977):
"Sequential procedures for detecting parameter changes in a time series model".
Journal of the American Statistical Association. Sept. 77, Vol.72, N°359, pp. 593-597.

[2] M. BASSEVILLE (1977):
"Ruptures de modèles statistiques".
Publication interne I.R.I.S.A. N° 91: "Traitement d'images".

[3a] M. BASSEVILLE, B. ESPIAU, J. GASNIER (1981):
"Edge detection using sequential methods for change in level - Part I: A sequential edge detection algorithm".
I.E.E.E. Trans. on A.S.S.P. Vol. ASSP-29, N°1, pp. 24-31; Feb. 1981.

[3b] M. BASSEVILLE (1981):
"Edge detection using sequential methods for change in level - Part II: Sequential detection of change in mean".
I.E.E.E. Trans. on A.S.S.P. Vol. ASSP-29, N°1, pp. 32-50; Feb. 1981.

[3c] M. BASSEVILLE, A. BENVENISTE (1981):
"An example of failure detection: design and comparative study of some algorithms".
Submitted to I.E.E.E. Trans. on A.C.

[4] T. BOHLIN (1977):
"Analysis of EEG signals with changing spectra using a short word Kalman estimator".
Math. Biosciences. Vol.35, pp. 221-259.

[5] T. BOJDECKI (1979):
"Probability maximizing approach to optimal stopping and its application to a disorder problem".
Stochastics. Vol.3, N°1, pp. 61-71.

[6] L.I. BORODKIN, V.V. MOTTL' (1976):
"Algorithm for finding the jump times of random process equation parameters".
Automatica i Telemekhanika. Vol.37, N°6, Part 1; pp. 23-32. June 1976.

[7] R.L. BROWN, J. DURBIN, J.M. EVANS (1975):
"Techniques for testing the constancy of regression relationships over time".
Jal. of the Royal Statistical Society (B). N°2, pp. 149-192.

[8] R. BUENO, E.Y. CHOW, K.P. DUNN, S.B. GERSHWIN, A.S. WILLSKY (1976):
"Status report on the generalized likelihood ratio failure detection technique, with application to the F8-aircraft".
Proc. IEEE Conf. on Decision and Control. Clearwater. Florida. Dec. 76; pp. 38-47.

[9] V.A. BUKHALEV (1976):
"Optimal filtering in systems with random step structure".
Automatica i Telemekhanika. Vol.37, N°2, Part 1; pp. 44-54. Feb. 76.

[10] B. CHALMOND (1979):
"Rupture de modèles pour des processus autorégressifs".
Thèse de 3ème cycle, Orsay. Nov. 79.

[11a] C.B. CHANG, K.P. DUNN (1978):
"A recursive generalized likelihood ratio test algorithm for detecting sudden changes in linear discrete systems".
Proc. IEEE Conf. on Decision and Control. San Diego. California. 1978; pp. 731-736.

[11b] C.B. CHANG, K.P. DUNN (1979):
"On GLR detection and estimation for unexpected inputs in linear discrete systems".
IEEE Trans. on Automatic Control. Vol. AC-24, N°3, pp. 499-501. June 79.

[12] T.T. CHIEN, M.B. ADAMS (1976):
"A sequential failure detection technique and its application".
IEEE Trans. on A.C. Vol. AC-21, N°5, pp. 750-757. Oct. 76.

[13] E.Y. CHOW (1980):
"Failure detection system design methodology".
Thesis. M.I.T. Oct. 1980.

[14a] B.S. DARKHOVSKII (1976):
"A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables".
Theory of Proba. and its Applications. Vol. 21, N° 1, pp. 178-183. March 76.

[14b] B.S. DARKHOVSKII, B.E. BRODSKII (1979):
"An identification of the "disorder" time of the random sequence".
Proc. 5th IFAC Symposium. Darmstadt. Sept. 79; pp. 373-379.

[15] M.H.A. DAVIS (1975):
"The application of nonlinear filtering to fault detection in linear systems".
IEEE Trans. on A.C. Vol. AC-20, N° 2, pp. 257-259. April 1975.

[16] J.C. DECKERT, M.N. DESAI, J.J. DEYST, A.S. WILLSKY (1977):
"F8 DFBW Sensor Failure Identification using Analytic Redundancy".
IEEE Trans. on A.C. Vol. AC-22, N°5, pp. 725-803. Oct. 77.

[17a] J. DESHAYES, D. PICARD (1979):
"Tests de rupture dans un modèle".
Comptes-rendus de l'Académie des Sciences, Paris, t.288 (12 mars 1979). Série A; pp. 563-566.

[17b] J. DESHAYES, D. PICARD (1979):
"Tests de rupture de régression: comparaison asymptotique".
Prépublication 79T23. Univ. de Paris-Sud. Orsay.

[18] B. ESPIAU, J.Y. CATROS (1980):
"Use of Optical Reflectance Sensors in Robotics Applications".
IEEE Trans. on Systems, Man, and Cybernetics. Dec. 1980. Vol. SMC-10, N°12.

[19] D.D. FALCONER, R.D. GITLIN (1978):
"Optimum reception of digital data signals in the presence of timing-phase hits".
Bell Syst. Techn. Jal. Vol.57, N°9, pp. 3181-3208. Nov. 78.

[20] B. FRIEDLAND (1979):
"Maximum likelihood estimation of a process with random transitions (failures)".
IEEE Trans. on A.C. Vol. AC-24, N°6, pp. 932-936. Dec. 79.

[21] K. GARBADE (1977):
"Two methods for examining the stability of regression coefficients".
Jal. of the American Stat. Assoc. Vol.72, N°357, pp. 54-63. March 77.

[22] J. GASNIER (1980):
"Détection et suivi de contours: étude comparative de méthodes séquentielles et application à des radiographies et des images T.V.".
Thèse de 3ème cycle. Rennes. Oct. 80.

[23] K. GIRIDHARAGOPAL, B. PAGUREK, C.M. WOODSIDE (1978):
"A modelling approach to state estimation in systems with switching parameters". IEEE Trans. on A.C. Vol. AC-23, N°5, pp. 948-950. Oct. 78.

[24a] A.H. GRAY, J.D. MARKEL (1976):
"Distance measures for Speech Processing".
IEEE Trans. on A.S.S.P. Vol. ASSP-24, N°5, pp. 380-391. Oct. 76.

[24b] R.M. GRAY, A. BUZO, A.H. GRAY, Y. MATSUYAMA (1980):
"Distorsion measures for Speech Processing".
IEEE Trans. on A.S.S.P. Vol. ASSP-28, N°4, pp. 367-376. Aug. 80.

[25a] D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER (1978):
"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part I: Identification of Persistent Rhythms".
IEEE Trans. on Biomedical Engineering. Vol. BME-25, N°4, pp. 344-353. July 78.

[25b] D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER (1978):
"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part II: Identification of Transient Rhythms".
IEEE BME-25. N°4, pp. 353-361.

[26] D.M. HAWKINS (1977):
"Testing a sequence of observations for a shift in location".
Jal of the American Stat. Assoc. Vol.72, N°357, pp. 180-186. March 77.

[27] W.G.S. HINES (1976):
"A simple monitor of a system with sudden parameter changes".
IEEE Trans. on Information Theory. Vol. IT-22, N°2, pp. 210-216. March 76.

[28] W.G.S. HINES (1976):
"Improving a simple monitor of a system with sudden parameter changes".
IEEE Trans. on I.T. Vol. IT-22, N°4, pp. 496-499. July 76.

[29] N. ISHII, A. IWATA, N. SUZUMURA (1979):
"Segmentation of nonstationary time-series".
Int. Jal of Systems Sciences. Vol.10, N°8, pp. 883-894. Aug. 79.

[30] N. ISHII, A. IWATA, N. SUZUMURA (1980):
"Detection of abrupt change and trend in the time-series".
Int. Jal of Systems Sciences. Vol.11, N°5, pp. 557-566. May 80.

[31] N. ISHII, H. SUGIMOTO, A. IWATA, N. SUZUMURA (1980):
"Computer classification of the EEG time-series by Kullback information measure".
Int. Jal of Systems Sciences. Vol.11, N°6, pp. 677-688. June 80.

[32] D. KAZAKOS, P. PAPANTONI-KAZAKOS (1980):
"Spectral distance measures between gaussian processes".
IEEE Trans. on A.C. Vol. AC-25, N°5, pp. 950-958. Oct. 80.

[33] T.H. KERR (1980):
"Statistical analysis of a two-ellipsoïd overlap test for real-time failure detection".
IEEE Trans. on Automatic Control. Vol. AC-25, N°4, pp. 762-772. Aug. 80.

[34] M. LABARRERE, M. PIRCHER, B. GIMONET, A. BUCHARLES, F. ESCAFFRE (1979):
"Recherche de méthodes de détection de pannes des capteurs".
Rapport final de convention DRET. ONERA-CERT. Toulouse.

[35a] S. LAKEHAL (1980):
"Aanalyse des signaux aléatoires non stationnaires par identification récursive".
Thèse de Docteur-Ingénieur ENST. Paris. July 80.

[35b] S. LAKEHAL (1980):
"On-line segmentation of quasi-stationary signals using autoregressive order selection".
First European Signal Processing Conference. Lausanne. 16-19 Sept. 80.

[36] A.F.S. LEE, S.M. HEGHINIAN (1977):
"A shift of the mean level in a sequence of independent normal random variables - A bayesian approach".
Technometrics. Vol.19, N°4, pp. 503-505. Nov. 77.

[37] V.Y. LUMEL'SKII (1972):
"An algorithm for detecting the instant of change of the properties of a random process". Automatica i Telemekhanika. Vol.10, pp. 67-73.

[38] B.P.M. Mac CABE, M.J. HARRISON (1980):
"Testing the constancy of regression relationships over time using least squares residuals".
Applied Statistics. Vol.29, N°2, pp. 142-148.

[39a] R.K. MEHRA, J. PESCHON (1971):
"An innovations approach to fault detection and diagnosis in dynamic systems".
Automatica. Vol.7, pp. 637-640.

[39b] M. MILLNERT (1980):
"An approach to recursive identification of abruptly changing systems".
Report LiTH. I-0349. Dept. of Electrical Engineering. Linköping University. Sweden.

[40] A.V. MOZGALEVSKII (1978):
"Technical diagnostics: Continuous Systems".
Automatica i Telemekhanika. Vol.39, N°1, Part 2, pp. 145-166, Jan. 78.

[41] P.M. NEWBOLD, Y-C. HO (1968):
"Detection of changes in the characteristics of a gauss-markov process".
IEEE Trans. on Aerospace and Electronic Systems. Vol. AES-4, N°5, pp. 707-718, Sept. 68.

[42] I.V. NIKIFOROV (1979):
"Cumulative sums for detection of changes in random process characteristics".
Automatica i Telemekhanika. Vol.40, N°2, Part 1, pp. 48-58, Feb. 79.

[43] A.N. PETTITT (1980):
"A simple cumulative sum type statistic for the change-point problem with zero-one observations".
Biometrika. Vol.67, N°1, pp. 79-84. Jan. 80.

[44] A.C. SANDERSON, J. SEGEN, E. RICHEY (1980):
"Hierarchical Modeling of EEG Signals".
IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. PAMI-2, N°5, pp. 405-414. Sept. 80.

[45] P. SANYAL, C.N. SHEN (1974):
"Bayes' decision rule for rapid detection and adaptive estimation scheme with space applications".
IEEE Trans. on A.C. Vol. AC-19, pp. 228-231. June 74.

[46] J. SEGEN, A.C. SANDERSON (1980):
"Detecting changes in a time-series".
IEEE Trans. on I.T. Vol. IT-26, N°2, pp. 249-255. March 80.

[47] P.K. SEN (1980):
"Asymptotic theory of some tests for a possible change in the regression slope occuring at an unknown time point".
Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete. Vol.52, pp. 203-218.

[48] S.A. SHABAN (1980):
"Change-point problem and two-phase regression: an annotated bibliography".
Int. Stat. Review. Vol. 48, N°1, pp. 83-93. March 80.

[49] R. STEELE, J. GOODMAN (1977):
"Detection and selective smoothing of transmission errors in linear PCM".
B.S.T.J. Vol.56, N°3, pp. 399-409. March 77.

[50] R. STEELE, D.J. GOODMAN, C.A. Mac GONEGAL (1979):
"A difference detection and correction scheme for combating DPCM transmission errors".
IEEE Trans. on Communications. Vol. COM-27, N°1, pp. 252-255. Jan. 79.

[51] D.D. SWORDER, V.G. ROBINSON (1973):
"Feedback regulators for jump parameter systems with state and control dependent transition rates".
IEEE Trans. on A.C. Vol. AC-18, N°4, pp. 355-360. Aug. 73.

[52a] L. TELKSNYS (1972):
"Determination of changes in the properties and recognition of random processes with a complicated structure".
Proc. 5th. IFAC Symposium. Paris, 1972.

[52b] L. TELKSNYS (1975):
"Determination of changes in the properties of random processes in the case of incomplete a priori data".
Statist. Pb. Upravlenija. Trudy Seminary Processy Optimal. Upravlenija V Seckcija Vyp 12; 1975; pp. 9-26. Vilnius.

[52c] L. TELKSNYS (1978):
"Determination of the most probable changes in the properties of multivariable dynamics sytems with unknown parameters".
Proc. IFAC Symposium. 1978; pp. 1807-1814.

[53] J.K. TUGNAIT, A.H. HADDAD (1979):
"A detection-estimation scheme for state estimation in switching environments".
Automatica. Vol.15, pp. 477-481.

[54] A.S. WILLSKY (1976):
"A survey of design methods for failure detection in dynamic systems".
Automatica. Vol.12; pp. 601-611.

[55] A.S. WILLSKY, J.J. DEYST, B.S. CRAWFORD (1975):
"Two self-test methods applied to an inertial system problem".
Jal of Spacecraft and Rockets. Vol.12, N°7, pp. 434-437. July 75.

[56] A.S. WILLSKY, H.L. JONES (1974):
"A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes".
Proc. IEEE Conf. on Decision and Control. Phoenix. Arizona. Nov. 74; pp. 846-853.

[57] A.S. WILLSKY, H.L. JONES (1976):
"A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems".
IEEE Trans. on A.C. Vol. AC-21, N°1, pp. 108-112. Feb. 76.

[58] A.S. WILLSKY, E.Y. CHOW, S.B. GERSHWIN, C.S. GREENE, P.K. HOUPT, A.L. KURKJIAN (1980):
"Dynamic model-based techniques for the detection of incidents on freeways".
IEEE Trans. on A.C. Vol. AC-25, N°3; pp. 347-360. June 80.

[59] B. WITTENMARK (1979):
"A two-level estimator for time varying parameters".
Automatica. Vol.15, pp. 85-89.

[60] K.J. WORSLEY (1979):
"On the likelihood ratio test for a shift in location of normal populations".
Jal of the American Statistical Assoc. Vol.74, N°366; pp.365-367. June 79.

[61] T. YOSHIMURA, K. WATANABE, K. KONISHI, T. SOEDA (1979):
"A sequential failure detection approach and the identification of failure parameters".
Int. Jal of Systems Sciences. Vol.10, N°7, pp. 827-836. July 79.

[62] S. ZACKS (1980):
"Numerical determination of the distributions of stopping variables associated with sequential procedures for detecting epochs of shift in distributions of discrete random variables".
Commun. Stat. - Simula. Computa. 1980. B9(1); pp. 1-80.